UNIVERSITY OF CAMBRIDGE

Caius and Gonville College

Machine Learning for Materials Science

A thesis submitted for the degree of
Doctor of Philosophy at the University of Cambridge

Author :
Bertrand Rouet-Leduc

Supervisor :
Prof. Sir Colin Humphreys

October 11, 2017

Abstract

Machine Learning for Materials Science

Machine learning is a branch of artificial intelligence that uses data to automatically build inferences and models designed to generalise and make predictions. In this thesis, the use of machine learning in materials science is explored, for two different problems: the optimisation of gallium nitride optoelectronic devices, and the prediction of material failure in the setting of laboratory earthquakes.

Light emitting diodes based on III-nitrides quantum wells have become ubiquitous as a light source, owing to their direct band-gap that covers UV, visible and infra-red light, and their very high quantum efficiency. This efficiency originates from most electronic transitions across the band-gap leading to the emission of a photon. At high currents however this efficiency sharply drops.

In chapters 3 and 4 simulations are shown to provide an explanation for experimental results, shedding a new light on this drop of efficiency. Chapter 3 provides a simple and yet accurate model that explains the experimentally observed beneficial effect that silicon doping has on light emitting diodes. Chapter 4 provides a model for the experimentally observed detrimental effect that certain V-shaped defects have on light emitting diodes. These results pave the way for the association of simulations to detailed multi-microscopy.

In the following chapters 5 to 7, it is shown that machine learning can leverage the use of device simulations, by replacing in a targeted and efficient way the very labour intensive tasks of making sure the numerical parameters of the simulations lead to convergence, and that the physical parameters reproduce experimental results. It is then shown that machine learning coupled with simulations can find optimal light emitting diodes structures, that have a greatly enhanced theoretical efficiency. These results demonstrate the power of machine learning for leveraging and automatising the exploration

of device structures in simulations.

Material failure is a very broad problem encountered in a variety of fields, ranging from engineering to Earth sciences. The phenomenon stems from complex and multi-scale physics, and failure experiments can provide a wealth of data that can be exploited by machine learning. In chapter 8 it is shown that by recording the acoustic waves emitted during the failure of a laboratory fault, an accurate predictive model can be built. The machine learning algorithm that is used retains the link with the physics of the experiment, and a new signal is thus discovered in the sound emitted by the fault. This new signal announces an upcoming laboratory earthquake, and is a signature of the stress state of the material. These results show that machine learning can help discover new signals in experiments where the amount of data is very large, and demonstrate a new method for the prediction of material failure.

Author : Bertrand Rouet-Leduc

Supervisor : Prof. Sir Colin Humphreys

Preface

This thesis is submitted to the University of Cambridge for the degree of Doctor of Philosophy. It is an account of the research that I have conducted under the supervision of Prof. Sir C. J. Humphreys in the Department of Materials Science and Metallurgy at the University of Cambridge between October 2013 and May 2017. It is my own work and contains nothing which is the outcome of work done in collaboration with others, except as declared in the Preface and specified in the text and acknowledgements. This dissertation is less than 60,000 words in length and has not been submitted to any other institution for a degree.

Bertrand Rouet-Leduc

May 2017

Acknowledgements

First and foremost, I wish to thank my supervisor, Professor Sir Colin Humphreys, for his guidance and his help, and for the freedom he gave me in defining my research project.

I wish to thank Dr. Christopher Ren, Dr. James Griffiths, and Dr. Siyuan Zhang for inviting me to collaborate with them, and providing me with the opportunity to model and simulate their fascinating experimental work. I also wish to thank Andrew Phillips for his valuable advice, and the rest of the gallium nitride group for exposing me to their ideas and research.

I have also been very fortunate to initiate a collaboration between the Cambridge gallium nitride group and the Los Alamos National Laboratory with Dr. Kipton Barros, Dr. Nick Lubbers, Dr. Turab Lookman, and Dr. Paul Johnson, a collaboration that resulted in a good proportion of the work presented in this thesis.

I am grateful to my fiancée Claudia Hulbert, for her continuous support, for helping me stay the course, and with whom I have had the luck to work, as she brought her machine learning expertise to the collaboration with LANL.

I acknowledge the Cambridge Trust, and the Gonville and Caius college for their financial support. I also acknowledge the EPSRC for funding my research done in Cambridge, and the Los Alamos National Laboratory (LANL) Laboratory Directed Research and Development (LDRD) for funding the collaboration with CNLS.

Last but not least I am thanking my parents Valérie and Frédéric for their constant support and encouragements through those many years of study.

Publications

B. Rouet-Leduc, K. Barros, T. Lookman, C. J. Humphreys, Optimisation of GaN LEDs and the reduction of efficiency droop using active machine learning, Sci. Rep. 6 (2016) 24862.

B. Rouet-Leduc, C. Hulbert, K. Barros, T. Lookman, C. J. Humphreys, Automatized convergence of optoelectronic simulations using active machine learning, (submitted).

B. Rouet-Leduc, C. Hulbert, K. Barros, C. J. Humphreys, Automatized parametrization of optoelectronic simulations using active machine learning, (in preparation).

B. Rouet-Leduc, C. Hulbert, N. Lubbers, K. Barros, C. J. Humphreys, P. A. Johnson, Machine learning predicts laboratory earthquakes, (submitted).

C. X. Ren, B. Rouet-Leduc, J. Griffiths, E. Bohacek, M. Wallace, P. Edwards, M. Hopkins, D. Allsopp, M. Kappers, R. Martin, et al., Analysis of defect-related inhomogeneous electroluminescence in InGaN/GaN QW LEDs, Superlattices Microstruct. 99 (2016) 118–124.

J. T. Griffiths, S. Zhang, B. Rouet-Leduc, W. Y. Fu, A. Bao, D. Zhu, D. J. Wallis, A. Howkins, I. Boyd, D. Stowe, et al., Nanocathodoluminescence reveals mitigation of the stark shift in InGaN quantum wells by Si doping, Nano Lett. 15 (11) (2015) 7639–7643.

Nomenclature

Acronyms

AE	Acoustic emission
APSYS	Advanced physical models of semiconductor devices
CCD	Charge-coupled device
CL	Cathodo-luminescence
EDX	Energy dispersive X-ray spectroscopy
EGO	Efficient global optimisation
EL	Electro-luminescence
EQE	External quantum efficiency
FWHM	Full width half maximum
GP	Gaussian process
HAADF	High angle annular dark field
IQE	Internal quantum efficiency
LED	Light emitting diode
ML	Machine learning
MOVPE	Metal organic vapour phase epitaxy
MQW	Multiple quantum wells
QB	Quantum barrier
QW	Quantum well
QCSE	Quantum confined Stark effect
RF	Random forest
STEM	Scanning transmission electron microscope

Roman symbols

A, B, C	Coefficients of the ABC model for efficiency droop
a, c	Wurtzite lattice constants
c_{nj}	Electron non-radiative capture coefficient with center j
c_{pj}	Hole non-radiative capture coefficient with center j
c_{pj}	Hole non-radiative capture coefficient with center j

C_p	Specific heat
D	Optical modes density
D_n	Electron diffusion constant
D_p	Hole diffusion constant
$E_{\mathrm{f}n}$	Electron quasi-Fermi energy
$E_{\mathrm{f}p}$	Hole quasi-Fermi energy
E_{HH}	Heavy hole valence band energy
E_{LH}	Light hole valence band energy
E_{CH}	Crystal-field split-off hole valence band energy
E_{g}	Energy band gap
E_{C}	Energy of the bottom edge of the conduction band
E_{V}	Energy of the top edge of the valence band
f_{A}	Degree of ionisation of acceptors
$1 - f_{\mathrm{D}}$	Degree of ionisation of donors
f_{tj}	Occupancy of the j^{th} trap level
G_n	Electron generation rate due to external excitation
G_p	Hole generation rate due to external excitation
H	Hamiltonian of Schrödinger's equation , Heat sources
$\hbar$	Reduced Planck constant
$\boldsymbol{J}_n$	Vector of the current flux density of electrons
$\boldsymbol{J}_p$	Vector of the current flux density of holes
k	Boltzmann constant
$\boldsymbol{k}$	Wave vector in reciprocal space
n	Electron concentration, band index, refractive index
N_{A}	Density of acceptors
N_{D}	Density of donors
N_{C}	Effective conduction band density of states
N_{V}	Effective valence band density of states
P^{sp}	Spontaneous polarisation value
P^{pz}	Piezoelectric polarisation value
$\boldsymbol{P}$	Polarisation vector
P	Laser excitation power
p	Hole concentration
$\boldsymbol{p}$	Quantum mechanical momentum operator
P_ϵ	Hydrostatic energy shift in the conduction band
q	Elementary charge
$R_{\ rm}$	Semi-empirical well trapping rate
R_n	Electron recombination rate

R_p	Hole recombination rate
R_{rad}	Radiative recombination rate
R_{Auger}	Auger recombination rate
R_n^{tj}	Electron non-radiative recombination due to center j
R_p^{tj}	Hole non-radiative recombination due to center j
R^2	Coefficient of determination
$\boldsymbol{r}$	Position in direct space
$\boldsymbol{S}_{\text{rad}}$	Electromagnetic field vector
T	Temperature
$u_{n,\boldsymbol{k}}$	Periodic part of a Bloch wave
v_{sn}	Electron saturation velocity
v_{sp}	Hole saturation velocity
V_e	Potential energy that delimits the quantum well

Greek symbols

Δ_{cr}	Crystal field energy splitting
Δ_{SO}	Spin-orbit energy splitting
$\mathcal{E}$	Electric field
ϵ	Permittivity of the medium
ϵ_{xx}, ϵ_{yy}, ϵ_{zz}	Strain tensor elements
η	Efficiency
κ	Thermal conductivity
μ_n	Electron mobility
μ_p	Hole mobility
μ	Expectancy
ρ	Charge density, material density, density of states
σ	Standard deviation
ϕ	Electric potential field
$\Phi_{n,\boldsymbol{k}}$	n^{th} wavefunction solution to Schrödinger's equation

Contents

Chapter 1

Introduction

1.1 Gallium nitride light emitting diodes

1.1.1 Solid state lighting

Lighting is the second largest end-use of electricity by the buildings sectors in the US [DOE, 2010a] [DOE, 2010b] and third in Europe[Bertoldi and Atanasiu, 2007]. It is the first end-use of electricity in the commercial sector in the US [DOE, 2010c]. Electric lighting consumes 19% of the total worldwide electricity production, which is as much as all the electricity produced by gas power plants and 15% more than that produced by nuclear or hydroelectric power plants [Waide et al., 2006]. As most of the electricity comes from burning coal or natural gas, this bewildering consumption of electricity makes lighting one of the first causes of greenhouse gas emissions, with an amount of CO_2 equivalent to 70% of what all the cars worldwide exhaust [Waide et al., 2006]. Unlike heating for instance, where the efficiency can reach 100%, lighting is particularly inefficient: incandescent light bulbs convert as little as 5% of the electricity they consume into light while compact fluorescent lamps (CFL) 20% at most. Lighting is one of the most inefficient uses of electricity, and very sizeable energy savings are to be made in this area.

In spite of the interesting gain in efficiency that CFL bring, their spreading is not desirable due to their mercury content. For instance, the replacement of all incandescent light bulbs with CFLs would bring 29t of mercury into U.S. houses [Humphreys, 2008]. Thus, the next generation of efficient and non-toxic home and office lighting is likely to come from inorganic light emitting diodes (LED) [Humphreys, 2008].

The simplest LED consists in a small piece of semiconducting material doped with impurities such that a p-n junction is created that delimits the

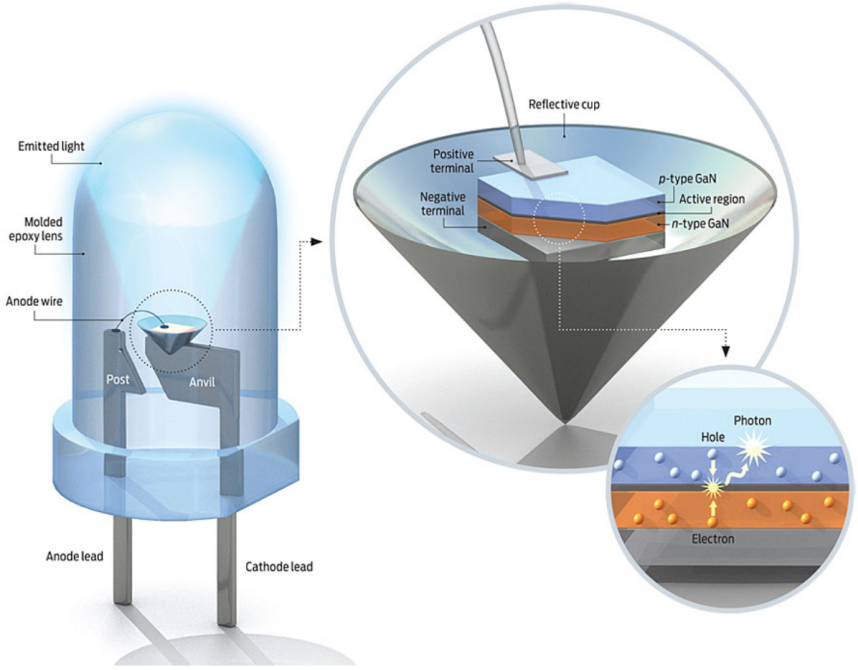


Figure 1.1: From [Stevenson, 2009]. Typical structure of an LED, consisting in a chip where light is produced out of recombining electron and holes, encapsulated to improve the extraction of light and dissipate the formation of heat.

p-side, where positively charged "holes" are abundant, from the n-side where negatively charged electrons are abundant. Under forward bias, where the p-side is connected to the positive terminal of the battery and the n-side is connected to the negative terminal of the battery, electrons and holes are pushed towards the p-n junctions where they can recombine (see Fig. 1.1). Upon recombination, light is emitted that carries the same energy as the energy separation between the electrons and the holes, the bandgap energy [Schubert et al., 2005]. The first visible LED was created in 1962 at the laboratories of General Electric by Holonyak and Bevacqua [Holonyak and Bevacqua, 1962] from a gallium arsenide-phosphide (GaAsP) p-n junction. Since the advent of this first red LED, emitting 0.1lm/W and that could only be seen in the dark, the efficacy has risen by a factor of 5 every 10 years, reaching over 100lm/W for today's red LEDs [Humphreys, 2008].

1.1.2 Gallium nitride light emitting diodes

Until the early 90's bright LEDs were restrained to a range of colours going from the infrared to the yellow. Gallium nitride-based LEDs were invented by Nakamura et al., [Nakamura et al., 1993] against the will of the company they were working for[Schubert et al., 2005]. To this end they used a very thin layer of indium gallium nitride (InGaN), sandwiched between gallium nitride (GaN) n and p sides, acting as a "quantum well" where electrons and holes are confined to recombine radiatively.

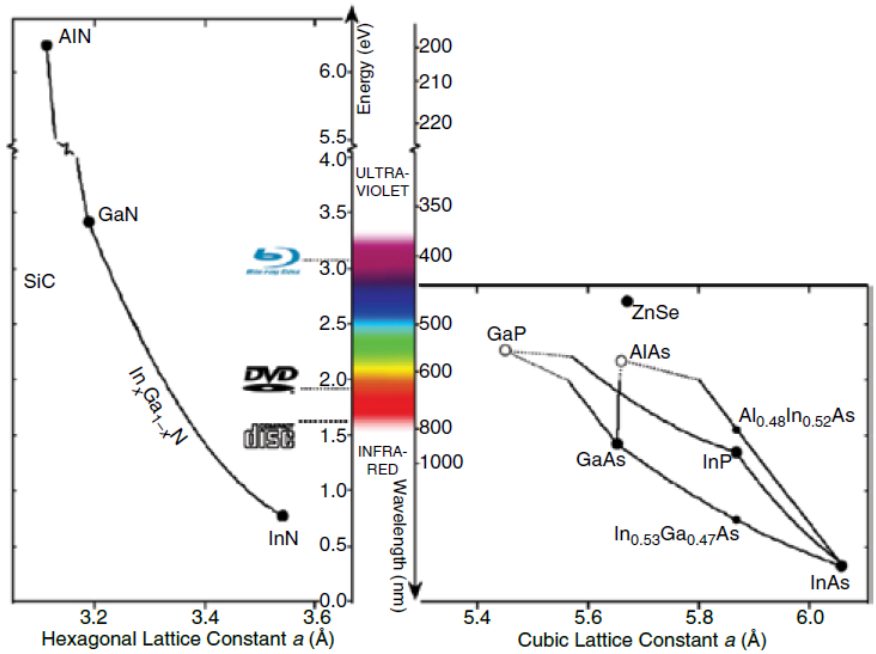


Figure 1.2: From [Humphreys, 2008]. Left: Bandgap energies of InN, GaN, AlN and their ternary alloys $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$. Right: Materials used for photo-emission before the invention of GaN-based LEDs.

As the bandgap of indium nitride (InN) is about 0.7eV, and that of GaN is of about 3.4eV, the mixing of the two, since both have a hexagonal crystal structure, makes it possible to engineer a bandgap that produces photons that span the full visible spectrum as well as infrared and near-UV light (see Fig. 1.2). It is now possible to emit bright light of any colour, although green GaN-based LEDs are less efficient than blue GaN-based LEDs, for reasons that are still debated [Humphreys, 2008]. White LEDs can be made in several ways: directly employing LEDs of different colours, such as red,

green, and blue enables to render a white spectrum, an example of which is having lower bandgap green and red active regions photo-excited by the emission of a biased higher blue bandgap active region [Schwach et al., 2007]. Coating a highly efficient blue or near ultra-violet LED with a phosphor that re-emits light of lower energy is also a common approach to design white LEDs [Sheu et al., 2003].

White LEDs already are more efficient than CFLs , the best conventional light sources, and unlike CFLs they hold the promise of yet more improvement for the years to come [Schubert et al., 2005] [Humphreys, 2008]. Thanks to Nakamura's discovery, gallium nitride can be considered among the most important new semiconductor materials since silicon [Humphreys, 2008] and within 20 years its applications have become ubiquitous, from LED backlighting applications, to computer, cellphone and TV LED screens, as well as private, commercial, and street lighting, representing an ever growing estimated market of $20B in 2014 [Haitz and Tsao, 2011], and $30B in 2016.

Nevertheless, as the current that passes through light emitting diodes is increased, their remarkable low-current efficiency quickly drops, reaching levels inferior to fluorescent tubes at very high currents. This efficiency "droop" has been referred to as the LED's "dark secret" [Stevenson, 2009]. It will be addressed in much more details in the next section (1.2).

1.2 LED efficiency droop

A perfect efficiency would be reached if every injected electron lead to the emission of one photon from the LED. Both electrons and photons can be lost in the process and the total quantum efficiency η_{EQE} is thus split into the internal quantum efficiency η_{IQE}, accounting for the loss of carriers, and the extraction efficiency η_{EXT} that accounts for the loss of photons that where successfully created within the LED:

$$\eta_{\mathrm{EQE}} = \eta_{\mathrm{IQE}} \times \eta_{\mathrm{EXT}} \tag{1.1}$$

It is generally admitted that the extraction efficiency η_{EXT} cannot explain the droop, although some extraction droop has been theorised for very thin LEDs [Bogdanov et al., 2010]. The internal quantum efficiency, defined as the fraction of the current that generates photons $\eta_{\mathrm{IQE}} = I_{\mathrm{rad}}/I$, is thus where efficiency droop lies. As of today, the physical origin of the efficiency droop is still highly controversial, with many different explanations having been proposed over the years. The following sections are an account of the most widespread explanations for droop at the time of writing.

1.2.1 Heating effects

Initially, heating effects were incriminated as the cause of droop, as they were found to be the cause of efficiency droop in gallium arsenide-based LEDs where thermal energy was enough to prevent the carriers from being trapped in their shallower quantum wells, but the deeper quantum wells of GaN-based LEDs are able to confine the carriers against thermal excitation much more effectively. Although heat has an important effect on GaN-based LEDs, the story doesn't end there as pulsed excitation conditions such that the structure can cool off between pulses doesn't prevent droop from happening [Efremov et al., 2006]. Furthermore, although the efficiency is vastly affected by the structure's temperature, its decrease with current is not, as the downward slope of efficiency with current from the peak efficiency does not change much with temperature [Kim et al., 2007] [Laubsch et al., 2009], eliminating heat as the cause of droop.

Although they are not explaining heating effects as the main cause for droop, Efremov and al. [Efremov et al., 2006] elucidated the effect of junction temperature on commercial LEDs efficiency. Comparing direct current with pulsed electro-luminescence (EL) measurements for which overheating is absent according to their thermal model, they could isolate part of the efficiency droop that is due to overheating. The authors also identified non-uniformity of the carrier injection as the cause of non-uniform heating and therefore non-uniform efficiency.

1.2.2 Auger recombination and the ABC model

The direct Auger process involves an electron-hole recombination leading to the excitation of another carrier instead of the emission of a photon [Delaney et al., 2009]. This other carrier can either be an electron (eeh processes) or a hole (hhe). As Auger processes involve three carriers their number scales as the cube of the carrier concentration and could explain the droop since the radiative recombination involves two carriers and thus scales only as the square of the carrier concentration. This cubed Auger term versus squared radiative term is related to the ABC model, commonly used to interpret the internal quantum efficiency of LEDs as a function of carrier concentration:

$$\eta_{\mathrm{IQE}}(n) = \frac{Bn^2}{An + Bn^2 + Cn^3}, \tag{1.2}$$

where the IQE is interpreted in terms of the ratio of carriers producing a photon (Bn^2) to those recombining non radiatively with defects (An) and those recombining non radiatively through Auger processes (Cn^3).

Direct Auger processes have been shown theoretically to be an unlikely explanation for the droop of GaN-based LEDs [Hader et al., 2008]. However, it has been shown recently that first principle calculations can derive Auger coefficients C high enough to account for droop, provided indirect Auger processes are considered [Kioupakis et al., 2011], where scattering mechanisms such as phonons and random alloy fluctuations provide additional degrees of freedom for the Auger processes to be quantum mechanically allowed and thus probable.

Experimental C Auger coefficients are usually retrieved through models that assume Auger to cause droop [Shen et al., 2007], thus yielding Auger coefficients high enough to cause droop, and successful fits of the ABC model cannot be seen as the confirmation of the underlying physical assumptions [Piprek, 2010]. On the other hand, it has been shown very recently that the detection of Auger electrons is proportional to the observed droop [Iveland et al., 2013]: Iveland and al. performed electron emission spectroscopy measurements on InGaN/GaN LEDs, recording the energy of electrons emitted from the device as a function of the current injected into it. They found that the onset of droop is correlated with the appearance of a high energy peak in the electron emission spectra. The integration of this high energy peak is linearly proportional to the droop current (that they defined as the additional current required for the device to produce a given light output compared to a hypothetical 100% efficiency device). The authors thus identified Auger recombination as the origin of droop. Nevertheless, their conclusion relied on the assumption (criticized in [Sarkissian et al., 2013]) that no other phenomenon could have caused high-energy electrons to be measured.

1.2.3 Carrier leakage, injection and distribution

The flow of electrons beyond the active region is commonly referred to as "carrier leakage" and is the reason for the implementation of an AlGaN "electron blocking layer" (EBL) on the p-side of the active region. Several papers [Pope et al., 2003] [Rozhansky and Zakheim, 2007] [Kim et al., 2007] [Schubert et al., 2009] [Han et al., 2009] [Liu et al., 2008b] have pointed to carrier leakage as a possible explanation for droop, arguing that in spite of the EBL a number of electrons escape from the active region and are thus lost with respect to radiative recombination.

Carrier leakage, experimentally demonstrated by measuring radiative recombination outside the active region on the p-side of a device [Vampola et al., 2009], has received a lot of attention and several explanations exist that are not mutually exclusive:

- Poor hole injection can be related to carrier leakage. Holes have a larger effective mass than electrons, *i.e.* their velocity over distances much larger than the lattice spacing as influenced by the crystal periodic potential is lower. As a result the concentration of holes in the active region is generally lower than the concentration of electrons [Xie et al., 2008] [Ni et al., 2008], a discrepancy that gets worse at high injection currents, leading to droop.

- polarisation mismatch is known to cause the heterointerfaces of the LED to be charged [Kim et al., 2007] [Schubert et al., 2008]. This charge accumulation has been shown to lower the p-side of the conduction band compared to the n-side, allowing the electrons to flow to the p-side outside the active region [Kim et al., 2007]. It has also been shown [Ryou et al., 2009] that the electric fields created by such charge accumulation are strong enough to confine electrons outside the active region against the EBL, accounting for carrier leakage.

- Schubert and al. [Schubert et al., 2007] proposed a modified ABC model that fits droop curves with a carrier leakage term instead of an Auger term. The usual Cn^3 Auger term is replaced by the difference between the experimental total efficiency η_{EQE} derived from pulsed EL measurements and the fit of the $An + Bn^2$ term of the canonical ABC model at low currents. The authors show that the droop term derived from this semi-fitting approach appears as a function depending on forward current rather than on carrier density, and remains very similar with respect to injected current for two MQW LEDs of very different dislocation density, consistent with carrier leakage as an explanation of droop.

A non-uniform distribution of carriers through the active region can be related to poor hole transport and injection. David and al. [David et al., 2008] showed that in conventional MQW LEDs most of the recombination events take place in the quantum well closest to the p layer. They reported far-field pattern electroluminescent measurements that barely change with a varying number of quantum wells, from low to high injection currents, and that can only be fitted by theoretical far-field pattern calculations from at most the two first quantum wells, with the best fit corresponding to only the first quantum well emitting.

1.2.4 Carrier localisation

In the recent years carrier localisation within the active region plane (in addition to carrier confinement perpendicular to the active region) has been considered as an explanation for the efficiency of GaN-based LEDs and thus a possible explanation for droop as well. The high efficiency of GaN-based LEDs at low currents is often seen as the fortunate result of carrier localisation [Davies et al., 2013] [Oliver et al., 2010] [Schomig et al., 2004] [Teo et al., 1998] or "anti-localisation" [Hangleiter et al., 2005] [Hangleiter et al., 2007], and a reduced localisation at high currents has been proposed as a cause of efficiency droop.

In order to demonstrate the existence of localised states, Schömig and al. [Schomig et al., 2004] studied an InGaN/GaN quantum well with sub-wavelength spatial resolution PL spectra at very low temperatures (4K), finding very sharp (0.8 meV) emission lines. The authors attributed these sharp peaks to the existence of localised quantum-dot-like states, due to their behaviour with rising temperature, that differs much with macroscopic PL measurements, as they undergo a blueshift compatible with a screening of the quantum confined Stark effect with rising temperature. Even more interestingly, they found that upon increasing the laser excitation power P single peaks do not shift while new higher energy peaks appear, explaining the blue-shift of macroscopic PL measurements with increasing excitation by the population of localised higher energy single exciton states (peak intensity $\propto$ P), as well as the formation of biexciton states (peak intensity $\propto$ P^2).

Chichibu and al. [Chichibu et al., 2006] demonstrated the importance of localisation in GaN-based LEDs using positron annihilation as an experimental hole recombination model, arguing that the very short positron localisation length observed in In-containing alloys proves a very short localisation length of holes as well.

Very recently, Davies and al. [Davies et al., 2013] reported time-resolved photoluminescence spectra of InGaN/GaN LEDs at low temperatures, finding that a fast-decaying component appears at high excitation powers, where droop occurs. On the basis of the concomitant disappearance of the S-shape dependence of the peak position with temperature observed by Hammersley and al. [Hammersley et al., 2012], a disappearance attributed itself to a saturation of localised states, Davies and al. attributed the fast-decaying emission of the PL spectra to weakly localized carriers.

1.3 Simulation of gallium nitride opto-electronic devices

The simulation of light emitting diodes consists in the numerical implementation of equations modelling their static (equilibrium) properties as well as their operation under electrical or optical excitation. It can be seen as a multi-scale simulation where classical "macroscopic" models are coupled with "microscopic" quantum models.

1.3.1 Classical models

Basic classical models [Sze and Ng, 2006], electrostatic equations, current-density equations and continuity equations form the drift-diffusion model, that accounts for the electrical behaviour of the devices.

Drift diffusion model

The Poisson equation (also known as Gauss's law and one of the Maxwell equations) relates electrical field to charge distribution [Sze and Ng, 2006]:

$$-\nabla^2\phi = \frac{\rho}{\epsilon} = \frac{q}{\epsilon}\left(-n + p + N_{\mathrm{D}}(1 - f_{\mathrm{D}}) - N_{\mathrm{A}} f_{\mathrm{A}} + \sum_j N_{tj}(\delta_i - f_{tj})\right), \tag{1.3}$$

with ϕ the electrical potential field, ϵ the permittivity of the medium, ρ the charge density, N_{A} and N_{D} the density of acceptors and donors, f_{A} and $(1 - f_{\mathrm{D}})$ the degree of ionisation of acceptor and donor levels, f_{tj} the occupancy of the j^{th} trap level given further in Eq. 1.20, and δ_i a constant set to 1 for donor-like traps and 0 for acceptor-like traps. The partial ionisation of impurities is usually described with Arrhenius-like laws:

$$f_{\mathrm{D}} = \frac{1}{1 + \exp\left((E_{\mathrm{D}} - E_{fn})/kT\right)} \tag{1.4}$$

$$f_{\mathrm{A}} = \frac{1}{1 + \exp\left((E_{\mathrm{A}} - E_{fn})/kT\right)} \tag{1.5}$$

Current-density equations account for the conduction of current under steady-state conditions, caused by the electric field and the gradient of carrier concentration [Sze and Ng, 2006]:

$$\boldsymbol{J}_n = q\mu_n n\mathcal{E} + qD_n\nabla n \tag{1.6}$$

$$\boldsymbol{J}_p = q\mu_p p\mathcal{E} - qD_p\nabla p, \tag{1.7}$$

with $\boldsymbol{J}_n$ and $\boldsymbol{J}_p$ the current flux densities for electrons and holes, μ_n and μ_p the electron and hole mobilities, D_n and D_p the electron and hole diffusion constants and $\mathcal{E}$ the electric field. At high fields $\mu_n\mathcal{E}$ and $\mu_p\mathcal{E}$ are saturated. The Canali model [Canali et al., 1975] is commonly used to account for this saturation:

$$\mu_n = \frac{\mu_{0n}}{(1 + (\mu_{0n}\mathcal{E}/v_{sn})^{\beta_n})^{1/\beta_n}} \tag{1.8}$$

$$\mu_p = \frac{\mu_{0p}}{(1 + (\mu_{0p}\mathcal{E}/v_{sp})^{\beta_p})^{1/\beta_p}}, \tag{1.9}$$

with μ_{0n} and μ_{0p} low field electron and hole mobilities, v_{sn} and v_{sp} electron and hole saturation velocities, and β_n and β_p constants of the model.

The bulk electron and hole concentrations are given by the integration of a square-root density of states with the Fermi-Dirac distribution, yielding [Sze and Ng, 2006] $\mathcal{F}_{1/2}$ the Fermi-Dirac integral of order 1/2:

$$n = N_{\mathrm{C}}\mathcal{F}_{1/2}\left(\frac{E_{fn} - E_{\mathrm{C}}}{kT}\right) \tag{1.10}$$

$$p = N_{\mathrm{V}}\mathcal{F}_{1/2}\left(\frac{E_{\mathrm{V}} - E_{fp}}{kT}\right), \tag{1.11}$$

with E_{fn} and E_{fp} the quasi-Fermi energies of electron and holes, N_{C} and N_{V} the effective density of states in the conduction and valence band, E_{C} the bottom edge of the conduction band, and E_{V} the top edge of the valence band.

The continuity equations account for time-dependant variations of carrier concentrations. The change over time of the carrier concentration and the concentration of ionised impurities is the difference between generation and recombination of carriers plus the current flow [Sze and Ng, 2006]:

$$\frac{\delta n}{\delta t} + N_{\mathrm{D}}\frac{\delta f_{\mathrm{D}}}{\delta t} = G_n - R_n + \frac{1}{q}\nabla \cdot \boldsymbol{J}_n \tag{1.12}$$

$$\frac{\delta p}{\delta t} - N_{\mathrm{A}}\frac{\delta f_{\mathrm{A}}}{\delta t} = G_p - R_p - \frac{1}{q}\nabla \cdot \boldsymbol{J}_p, \tag{1.13}$$

where G_n and G_p are the electron and hole generation rates due to external excitation, R_n and R_p electron and hole recombination rates accounting for various phenomena that tend to restore the equilibrium.

These phenomena [Schubert et al., 2005] include radiative recombination, induced by energy conservation of an electron falling back into the valence band from the conduction band, Auger recombination that describes the

transfer of energy to another electron or hole, and recombination via defects and other carrier traps.

Radiative recombination can be expressed as:

$$R_{\text{rad}} = Bnp = B\left(n_0 + \Delta n(t)\right)\left(p_0 + \Delta p(t)\right), \tag{1.14}$$

which simplifies for high-level excitations ($\Delta n(t) \gg n_0, p_0$) to:

$$R_{\text{rad}} = B\Delta n(t)^2, \tag{1.15}$$

with n and p the electron and hole concentrations, n_0 and p_0 the equilibrium carrier concentrations, $\Delta n(t) = \Delta p(t)$ the excess electron and hole concentrations, and B the "bimolecular" recombination coefficient.

Auger mechanisms (See Chap. 1.2.2) describe how the energy dissipated by the recombination between an electron and a hole (approximately the band gap) can be given to an other carrier, electron or hole, highly excited without moving to an other energy band, thus losing its energy through phonon emissions. Suspected to play an important part at high-level excitations [Iveland et al., 2013], it can then be expressed as:

$$R_{\text{Auger}} = Cn(t)^3, \tag{1.16}$$

with C the Auger coefficient [Olshansky et al., 1984].

Recombination via defects is the most common cause for non-radiative recombination, and includes interaction with foreign atoms, point defects, dislocations and combinations of them [Longini and Greene, 1956]. These defects are associated with energy levels that can exist within the energy gap of the semiconductor, thus providing an efficient recombination center, known as a "luminescence killer". The non-radiative recombination rates caused by such a center j are given by [Shockley and Read, 1952]:

$$R_n^{tj} = c_{nj} n N_{tj}(1 - f_{tj}) - c_{nj} n_{1j} N_{tj} f_{tj} \tag{1.17}$$

$$R_p^{tj} = c_{pj} p N_{tj} f_{tj} - c_{pj} p_{1j} N_{tj}(1 - f_{tj}), \tag{1.18}$$

with the following trap dynamics equation holding [Smith and Smith, 1978]:

$$N_{tj}\frac{\delta f_{tj}}{\delta t} = R_n^{tj} - R_p^{tj}, \tag{1.19}$$

with c_{nj} and c_{pj} electron and hole capture coefficients that relate their lifetime to the concentration of recombination centers: $\frac{1}{\tau_{nj}} = c_{nj} N_{tj}$ and $\frac{1}{\tau_{pj}} = c_{pj} N_{tj}$, n_{1j} and p_{1j} the electron and hole concentrations if the Fermi energy was

located at the trap level. Equations 1.17, 1.18, 1.19 yield the occupancy of a trap j used in Eq. 1.3:

$$f_{tj} = \frac{c_{nj} n + c_{pj} p_{1j}}{c_{nj}(n + n_{1j}) + c_{pj}(p + p_{1j})} \tag{1.20}$$

Also relevant to the simulation of carrier recombination is the non-radiative recombination at surfaces: as the surface rearranges itself due to the ending of the periodic structure of the semi-conductor, different energy states arise that can be located within the energy gap, thus acting as efficient recombination centers. These effects are usually accounted for with phenomenological models, and a very high density N_{tj} of trap states near the surface of the device can be used in simulations to reproduce the substantial recombination that can occur at the surface.

Thermal effects

Heating effects have an important impact on the behaviour of semi-conductor devices and LEDs in particular [Efremov et al., 2006]. Hence, heat flow and temperature need to be modelled as well and evaluated on the device being simulated, the temperature being a parameter of most of the equations modelling the operation of an LED. Heat propagation is modelled through the usual heat equation:

$$C_p \rho \frac{\delta T}{\delta t} = \kappa \Delta T + H, \tag{1.21}$$

with C_p the specific heat, ρ the material's density, κ the thermal conductivity, and H the heat sources within the material, as follows [Wachutka, 1990]:

$$\begin{aligned} H &= q\frac{J_n^2}{\mu_n n} + q\frac{J_p^2}{\mu_p p} \\ &+ q(R-G)\left(E_{fp} - E_{fn} + T\left((\delta E_{fn}/\delta T)_{n,p} - (\delta E_{fp}/\delta T)_{n,p}\right)\right) \\ &+ qT\left((\delta E_{fn}/\delta T)_{n,p} - P_n\right) \operatorname{div} \boldsymbol{J}_n \\ &- qT\left((\delta E_{fp}/\delta T)_{n,p} + P_p\right) \operatorname{div} \boldsymbol{J}_p \\ &- qT(\boldsymbol{J}_n \cdot \nabla P_n + \boldsymbol{J}_p \cdot \nabla P_p) - \langle \operatorname{div} \boldsymbol{S}_{\text{rad}} \rangle, \end{aligned} \tag{1.22}$$

with the first line accounting for the Joule heat of electrons and holes, the second line accounting for the heat due to recombination of electrons and holes, the third line accounting for the Thomson heat exchanged between the lattice and the carriers as they flow through regions of varying thermoelectric powers [Stratton, 1972] P_n and P_p (due to varying temperature or doping

concentrations), and the fourth line accounting for the local absorption of radiative energy into the materials from an electromagnetic field $\boldsymbol{S}_{\text{rad}}$.

Additionally, a thermal gradient induces a current that supplements the electron current by: [Wachutka, 1990]:

$$-q\mu_n n P_n \nabla T, \tag{1.23}$$

and supplements the hole current by:

$$-q\mu_p p P_p \nabla T, \tag{1.24}$$

1.3.2 Quantum mechanical models

The theories described so far have dealt with the description of bulk material, and the main region of interest of an LED, its active region, remains to be modelled. The active region of an LED is by essence a heterostructure (i.e. a composite of materials) with quantum effects, and its modelling requires to heavily rely upon quantum mechanical theory. First-principles calculations of the electronic structure of wurtzite-type GaN exist [Suzuki et al., 1995], but their computational cost prohibit their use in device-wide calculations where "simpler" approximations must be used, such as the $\mathrm{k}\cdot\mathrm{p}$ theory.

The $\mathrm{k}\cdot\mathrm{p}$ theory

The $\mathrm{k}\cdot\mathrm{p}$ theory [Luttinger and Kohn, 1955] can be very briefly explained as a perturbation theory applied to stationary Schrödinger-type equations of periodic crystals:

$$H\,\Phi_{n,\boldsymbol{k}} = E_{n,\boldsymbol{k}}\,\Phi_{n,\boldsymbol{k}}, \tag{1.25}$$

with H the Hamiltonian of the system, Φ and E a couple eigenvector-eigenvalue solution to the equation, n a discrete parameter counting the solutions, and the crystal wave vector $\boldsymbol{k}$ a continuous parameter describing the solution in reciprocal space, and with Φ a Bloch wave:

$$\Phi_{n,\boldsymbol{k}}(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n,\boldsymbol{k}}(\boldsymbol{r}), \tag{1.26}$$

with $\boldsymbol{r}$ the position in direct space and u a function that has the same periodicity as the crystal. In case of a non-degenerate band and without considering spin-orbit coupling, Eq. 1.25 is equivalent to:

$$\left(H + \hbar\frac{\boldsymbol{k}\cdot\boldsymbol{p}}{m}\right) u_{n,\boldsymbol{k}} = \left(E_{n,\boldsymbol{k}} - \frac{\hbar^2 k^2}{2m}\right) u_{n,\boldsymbol{k}}, \tag{1.27}$$

with $\boldsymbol{p} = -i\hbar\boldsymbol{\nabla}$. The term $\hbar\frac{\boldsymbol{k}\cdot\boldsymbol{p}}{m}$ is then treated as a perturbation of H, which yields:

$$E_{n,\boldsymbol{k}} - \frac{\hbar^2 k^2}{2m} = E_{n,0} + \langle u_{n,0}|\hbar\frac{\boldsymbol{k}\cdot\boldsymbol{p}}{m}|u_{n,0}\rangle + \frac{\hbar^2}{m^2}\sum_{n'\neq n}\frac{|\langle u_{n',0}|\boldsymbol{k}\cdot\boldsymbol{p}|u_{n,0}\rangle|^2}{E_{n,0}-E_{n',0}} + O(\|\boldsymbol{k}\|^3) \tag{1.28}$$

Chuang and Chang have applied the $\mathrm{k}\cdot\mathrm{p}$ theory to wurtzite strained crystals [Chuang and Chang, 1996] taking into account spin-orbit coupling and using perturbation theory for degenerate bands, yielding the following six-by-six Hamiltonian in second order of $\boldsymbol{k}\cdot\boldsymbol{p}$:

$$H = \begin{pmatrix} F & -K_k^* & -H_k^* & 0 & 0 & 0 \\ -K_k & G & H_k & 0 & 0 & \Delta \\ -H_k & H_k^* & \lambda & 0 & \Delta & 0 \\ 0 & 0 & 0 & F & -K_k & H_k \\ 0 & 0 & \Delta & -K_k^* & G & -H_k^* \\ 0 & \Delta & 0 & H_k^* & -H_k & \lambda \end{pmatrix}, \tag{1.29}$$

developped on the following basis vectors:

$$\begin{aligned} |u_1\rangle &= \frac{-1}{\sqrt{2}}|(X+iY)\uparrow\rangle \\ |u_2\rangle &= \frac{1}{\sqrt{2}}|(X-iY)\uparrow\rangle \\ |u_3\rangle &= |Z\uparrow\rangle \\ |u_4\rangle &= \frac{1}{\sqrt{2}}|(X-iY)\downarrow\rangle \\ |u_5\rangle &= \frac{-1}{\sqrt{2}}|(X+iY)\downarrow\rangle \\ |u_6\rangle &= |Z\downarrow\rangle \end{aligned} \tag{1.30}$$

with X, Y and Z functions that have the same symmetry as the atomic functions p_x, p_y and p_z in the wurtzite structure, and with the matrix elements

given by:

$$\begin{aligned}
F &= \Delta_1 + \Delta_2 + \lambda + \theta \\
G &= \Delta_1 - \Delta_2 + \lambda + \theta \\
\lambda &= \frac{\hbar^2}{2m_0}(A_1 k_z^2 + A_2(k_x^2 + k_y^2)) + \lambda_\epsilon \\
\lambda_\epsilon &= D_1\epsilon_{zz} + D_2(\epsilon_{xx} + \epsilon_{yy}) \\
\theta &= \frac{\hbar^2}{2m_0}(A_3 k_z^2 + A_4(k_x^2 + k_y^2)) + \theta_\epsilon \\
\theta_\epsilon &= D_3\epsilon_{zz} + D_4(\epsilon_{xx} + \epsilon_{yy}) \\
K_k &= \frac{\hbar^2}{2m_0}A_5(k_x + ik_y)^2 + D_5(\epsilon_{xx} + 2i\epsilon_{xy} - \epsilon_{yy}) \\
H_k &= \frac{\hbar^2}{2m_0}A_6(k_x + ik_y)k_z + D_6(\epsilon_{zx} + i\epsilon_{yz}) \\
\Delta &= \sqrt{2}\Delta_3,
\end{aligned} \tag{1.31}$$

where the $A_{i=1\ldots6}$ and $D_{i=1\ldots6}$ are band-edge parameters, $\Delta_1 = \Delta_{\text{cr}}$ with Δ_{cr} the crystal field splitting, $\Delta_2 = \Delta_3 = \frac{1}{3}\Delta_{\text{SO}}$ with Δ_{SO} the spin-orbit splitting, and where the elements of the strain tensor $\bar{\bar{\epsilon}}$ are the following for a strained layer of wurtzite crystal grown following the (0001) direction [Love, 1944]:

$$\begin{aligned}
\epsilon_{xx} &= \epsilon_{yy} = \frac{a_0 - a}{a} \\
\epsilon_{zz} &= -2\frac{C_{13}}{C_{33}}\epsilon_{xx},
\end{aligned} \tag{1.32}$$

with a_0 and a the lattice constants of the substrate and the grown layer respectively, C_{13} and C_{33} are the stiffness constants.

The valence bands (commonly labelled as heavy hole HH, light-hole LH, and crystal-field split-off hole CH) can then be expressed analytically [Chuang and Chang, 1996]. Near the zone center ($k \approx 0$) in second order of k:

$$\begin{aligned}
E_{\text{HH}} &= F \\
E_{\text{LH}} &= \frac{G+\lambda}{2} + \sqrt{\left(\frac{G-\lambda}{2}\right) + \Delta^2} \\
E_{\text{CH}} &= \frac{G+\lambda}{2} - \sqrt{\left(\frac{G-\lambda}{2}\right) + \Delta^2}
\end{aligned} \tag{1.33}$$

Away from the zone center (large k), the spin-orbit coupling can be neglected and the Hamiltonian has analytical solutions that yield the following

bands:

$$\begin{aligned} E_{\mathrm{HH}} &= F' - K_k \\ E_{\mathrm{LH}} &= \frac{1}{2}\left(F' + K_k + \lambda + \sqrt{(F' + K_k - \lambda)^2 + 8H_k^2}\right) \\ E_{\mathrm{CH}} &= \frac{1}{2}\left(F' + K_k + \lambda - \sqrt{(F' + K_k - \lambda)^2 + 8H_k^2}\right), \end{aligned} \tag{1.34}$$

with $F' = \Delta_1 + \lambda + \theta$.

The conduction-band edge is given as a function of strain by [Chuang and Chang, 1996]:

$$E_{\mathrm{C}} = E_{\mathrm{V}} + \Delta_1 + \Delta_2 + E_g + P_\epsilon, \tag{1.35}$$

with E_{V} the reference valence band-edge, E_g the energy gap, and $P_\epsilon = a_{\mathrm{cz}}\epsilon_{zz} + a_{\mathrm{ct}}(\epsilon_{xx} + \epsilon_{yy})$ the hydrostatic energy shift, with a_{cz} and a_{ct} the longitudinal and transverse deformation potentials.

The model parameters for bulk binary alloys GaN, AlN and InN can be determined experimentally as well as from first principle calculations [Vurgaftman et al., 2001], and that of ternary alloys can be fitted using a quadratic form:

$$E_g(A_{1-x}B_x) = (1-x)E_g(A) + xE_g(B) - x(1-x)C, \tag{1.36}$$

where C is the so-called bowing parameter that accounts for the deviation from Vegard's law.

Quantum wells

The previous section dealt with the derivation of the bulk band structure of strained wurtzite material, which is the basis for the approximation of the quantum states within strained GaN/InGaN quantum wells. The InGaN region of an GaN/InGaN quantum well acts as a confinement area. Its energy depth is given by the bulk energy levels of the materials, as given by the considerations from the previous section. Its confined energy levels are given by another Schrödinger-type equation accounting for the confined quantum states that exist within the well [Chuang and Chang, 1997]. For electrons:

$$H^c(k_t, k_z) = \frac{\hbar^2}{2}\left(\frac{k_t^2}{m_e^t} + \frac{k_z^2}{m_e^z}\right) + V_e(z) + P_\epsilon, \tag{1.37}$$

with $k_t = -i\nabla_t$, $k_z = -i\frac{\delta}{\delta z}$, m_e^z and m_e^t the electron effective masses following directions longitudinal and transverse with respect to the c axis, and $V_e(z)$ the potential energy formed by the unstrained conduction-band edge in the

previous section, Eq. 1.35, changing with z depending on the layer's material. Solving:

$$H^c(k_t = 0)\,\Phi_n(z) = E_n^c\,\Phi_n(z), \tag{1.38}$$

yields the conduction subbands:

$$E_n^c(\boldsymbol{k}) \approx E_n^c(k_t = 0) + \frac{\hbar^2 k_t^2}{2m_{e,w}^t}, \tag{1.39}$$

with $m_{e,w}^t$ the effective mass of an electron in the well following directions parallel to it, using the parabolic approximation in the plane of the quantum well (i.e. modelling it as an infinite plane), that describes the energy states available to electrons confined in the quantum well, with Φ_n the envelope function of the electron wave-function $\Psi_{n,k_t}^{c\eta}$, i.e. its periodic part in Bloch's expression (Eq. 1.28):

$$\Psi_{n,k_t}^{c\eta}(z) = \frac{e^{i\boldsymbol{k}_t \cdot \boldsymbol{r}_t}}{\sqrt{A}}\,\Phi_n(z)|S,\eta\rangle, \tag{1.40}$$

with S a function with the same symmetry in the wurtzite structure as the s atomic function, η the spin ($\uparrow$ or $\downarrow$), and A a normalisation factor equal to the area of the quantum well in the xoy plane transversal to the growth direction.

The derivation of the hole subbands is trickier as the hole bands (HH, LH and CH) are "mixed" by the Hamiltonian (i.e. its solutions are linear combinations of heavy hole, light hole, and crystal-field hole states). The Hamiltonian giving the approximation of the valence band structure in the previous section, Eq. 1.29, can be block-diagonalised [Chuang and Chang, 1996]:

$$H = \begin{pmatrix} H^U & 0 \\ 0 & H^L \end{pmatrix}, \tag{1.41}$$

with H^U and H^L 3×3 matrices, in a new basis $\{|1\rangle ... |6\rangle\}$ of mixed spin up and spin down states (see appendix for more details). In this new basis the hole wavefunctions, abiding to Bloch's theorem, can be expressed as [Chuang and Chang, 1996]:

$$\begin{aligned} \Psi_m^U(z;k_t) &= \frac{e^{i\boldsymbol{k}_t \cdot \boldsymbol{r}_t}}{\sqrt{A}}(u_m^1(z;k_t)|1\rangle + u_m^2(z;k_t)|2\rangle + u_m^3(z;k_t)|3\rangle) \\ \Psi_m^L(z;k_t) &= \frac{e^{i\boldsymbol{k}_t \cdot \boldsymbol{r}_t}}{\sqrt{A}}(u_m^4(z;k_t)|4\rangle + u_m^5(z;k_t)|5\rangle + u_m^6(z;k_t)|6\rangle), \end{aligned} \tag{1.42}$$

with $\Psi_m = \begin{pmatrix} \Psi_m^U \\ \Psi_m^L \end{pmatrix}$, with m denoting the valence subband index, and u_m^i the envelope periodic functions. The valence subbands are then given by solving [Chuang and Chang, 1996]:

$$\begin{aligned} (H^U + E_{\mathrm{V}}^0(z) I_3) \begin{pmatrix} u_m^1 \\ u_m^2 \\ u_m^3 \end{pmatrix} &= E_m^U(k_t) \begin{pmatrix} u_m^1 \\ u_m^2 \\ u_m^3 \end{pmatrix} \\ (H^L + E_{\mathrm{V}}^0(z) I_3) \begin{pmatrix} u_m^4 \\ u_m^5 \\ u_m^6 \end{pmatrix} &= E_m^L(k_t) \begin{pmatrix} u_m^4 \\ u_m^5 \\ u_m^6 \end{pmatrix}, \end{aligned} \tag{1.43}$$

with I_3 the 3×3 identity matrix.

The theory presented in these sections is aimed at enabling simulations at the device scale. Due to the very large scale this involves compared to purely *ab-initio* quantum calculations, a number of empirical approximations are required. Most notably, in order to be able to self-consistently solve the Poisson equations and the Schrödinger equations in a reasonable amount of time, the active region must be modelled fairly simply. Hence, recent discoveries that have shown the importance of local fluctuations in the composition and width of the quantum wells in localising carriers [Watson-Parris et al., 2011, Schulz et al., 2015, Humphreys et al., 2017] must be considered empirically, and cannot be simulated explicitly.

A fully atomistic model of an LED device is not possible from a computational standpoint. The overwhelming difference in scale between the Poisson equations that deal with currents through the whole device, and Schrödinger equations that deal with the carriers in the quantum wells is treated empirically, with the wells acting as traps for the flowing classical carriers, that then become localised and populate the quantum states of the wells [Alam et al., 2000]. The well trapping rate is given by the transition probabilities between 3D states in the barriers and 2D states in the wells:

$$R_{\mathrm{qw}} = \int \int g_{3\mathrm{D}}(E_{2\mathrm{D}}) g_{3\mathrm{D}}(E_{3\mathrm{D}}) S(E_{2\mathrm{D}}, E_{3\mathrm{D}}) \times f_{3\mathrm{D}}(E_{3\mathrm{D}})(1 - f_{2\mathrm{D}}(E_{2\mathrm{D}}))\mathrm{d}E_{2\mathrm{D}}\mathrm{d}E_{3\mathrm{D}} \tag{1.44}$$

with S the transition probability between 2D and 3D states, $g_{2\mathrm{D}}$ and $g_{3\mathrm{D}}$ the density of states, $E_{2\mathrm{D}}$ and E_{3D} their energy, and $f_{2\mathrm{D}}$ and $f_{3\mathrm{D}}$ the Fermi functions of the 2D and 3D states. Taking some approximations make this trapping simpler to model: ignoring the dependence of S on $E_{3\mathrm{D}}$, and considering that most confined carriers populate the bottom state of energy E_{QW} of the quantum well, the quantum well trapping rate reduces to [Alam et al.,

2000]:

$$R_{\mathrm{qw}} = \left(\frac{n_{\mathrm{3D}} - n_{\mathrm{3D,0}}}{\tau} \right) (1 - f_{\mathrm{2D}}(E_{\mathrm{QW}})) \tag{1.45}$$

with n_{3D} the barrier carrier concentration, $n_{3D,0}$ the barrier concentration at equilibrium, and τ the time constant associated with the trapping of carriers in the wells that empirically accounts for any phenomenon localising the carriers in the wells.

For the same reason of scale, defects in the quantum wells have to be treated empirically in Poisson-Schrödinger simulations. Much of the success of gallium nitride in optoelectronics is due to its resilience to defects. Several phenomena have been shown to explain this resilience, including V-shaped defects that inhibit the interaction between carriers and threading dislocations [Hangleiter et al., 2005]; as well as fluctuations in both the width and the alloy composition of the InGaN wells, localising further the carriers not only perpendicular to the well (keeping the carriers confined), but laterally as well, preventing the carriers from interacting with adjacent defects. A wealth of evidence [Humphreys et al., 2017] has shown that well width fluctuations (of atomic height and a few nm in diameter) are localising the electrons on the top (p-side) well surface, and that the random alloy fluctuations of the InGaN are localising the holes on the bottom (n-side) well surface. These mechanisms happen at a lateral scale of a few nm, while the simulation is device-wide, up to the scale of a few mm, and are treated empirically, both in the trapping rate of the carriers in the quantum wells (Eq. 1.3.2), and in the trapping rate of non-radiative recombination centers (Eqs. 1.17 and 1.18).

Light emission

The density of states, a measure of how many quantum states are available at a given energy, is the starting point for the computation of the light emission. For quantum wells it is directly derived from the subband structure [Chuang and Chang, 1997]:

$$\rho(E) = \frac{1}{2\pi} \sum_{m} \sum_{\substack{k \\ E_m(k)=E}} k \left| \frac{\mathrm{d}E_m}{\mathrm{d}k} \right|^{-1}, \tag{1.46}$$

where the first sum runs on all the subbands of the considered carriers, and the second sum runs on all the k such that the energy of the subband $E_m(k)$ is equal to the energy E where the density of states ρ is evaluated. The usual

2D density of states is step like when derived from simple models [Sze and Ng, 2006], but more complicated when band mixing effects are taken into account through the k · p method previously presented.

The spontaneous emission rate per unit energy at a given energy $R_{\mathrm{sp}}(E)$ is then derived from Fermi's Golden rule to yield [Yan et al., 1990]:

$$R_{\mathrm{sp}}(E) = \sum_{i,j} \left(\frac{2\pi}{\hbar}\right) \left(\frac{e}{m_0}\right)^2 \left(\frac{2\hbar\omega/\epsilon}{4\omega^2}\right) |M_{ij}|^2 \rho_{\mathrm{red}}^{ij}(E) D(E) f_i(1-f_j), \quad (1.47)$$

with the sum running over all allowed $i \rightarrow j$ transitions with i a conduction state and j a valence state, ϵ the optical dielectric constant, and $D(E)$ the optical mode density, given for a material of refractive index n by:

$$D(E) = \frac{n^3 E^2}{\pi^2 \hbar^3 c^3}, \quad (1.48)$$

and where ρ_{red}^{ij} is the density of states for which a reduced effective mass is used: $m_r = m_e m_h/(m_e + m_h)$. $M_{ij} = \langle i|\boldsymbol{p}|j\rangle$ is the momentum matrix element between the i and j states, and f_i and f_j are the Fermi functions of the i and j states. The global spontaneous radiative recombination rate is then the integration of this rate over all energies:

$$R_{\mathrm{rad}} = \int R_{\mathrm{sp}}(E)\mathrm{d}E, \quad (1.49)$$

1.3.3 Additional models specific to indium gallium nitride quantum wells

polarisation and quantum confined Stark effect

The piezoelectric nature of the wurtzite structure, along with the strain that exists in the LED structure due to the lattice mismatch between the layers, result in a large piezoelectric polarisation. Additionally, III-V wurtzite nitrides have an asymmetric bonding structure causing a spontaneous polarisation [Bernardini et al., 1997]. Both polarisations lead to large two-dimensional charges at the interfaces between nitride layers which in turn result in electrostatic fields and free-charge accumulation, provided the density of free-charge from doping is high enough [Bernardini and Fiorentini, 2001]. Fiorentini and al. [Fiorentini et al., 2002] have shown with comparisons between *ab initio* simulations and experiments that an accurate modelling of the polarisation of III-V ternary nitride alloys requires to account for non-linearities in the

spontaneous polarisation with respect to the values determined for the parent binary compounds (in C/m^2):

$$\begin{aligned} P^{\mathrm{sp}}_{\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N}} &= -0.090x - 0.034(1-x) + 0.019x(1-x) \\ P^{\mathrm{sp}}_{\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}} &= -0.042x - 0.034(1-x) + 0.038x(1-x) \\ P^{\mathrm{sp}}_{\mathrm{Al}_x\mathrm{In}_{1-x}\mathrm{N}} &= -0.090x - 0.042(1-x) + 0.071x(1-x) \end{aligned} \tag{1.50}$$

On the other hand, they showed [Bernardini and Fiorentini, 2001] that Vegard's law (linear interpolation) holds for the piezoelectric polarisation:

$$P^{\mathrm{pz}}_{\mathrm{X}_x\mathrm{Y}_{(1-x)}}(x) = xP^{\mathrm{pz}}_{\mathrm{XN}}(\epsilon(x)) + (1-x)P^{\mathrm{pz}}_{\mathrm{YN}}(\epsilon(x)), \tag{1.51}$$

with $\epsilon(x)$ the basal strain due to a lattice mismatch between the substrate and the alloy layer:

$$\epsilon(x) = \frac{a_{\mathrm{subs}} - a(x)}{a(x)}, \tag{1.52}$$

where the alloy lattice itself can also be derived from Vegard's law [Bernardini and Fiorentini, 2001]:

$$\begin{aligned} a_{\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N}}(x) &= 0.31986 - 0.00891x \\ a_{\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}}(x) &= 0.31986 + 0.03862x \\ a_{\mathrm{Al}_x\mathrm{In}_{1-x}\mathrm{N}}(x) &= 0.35848 - 0.04753x \end{aligned} \tag{1.53}$$

As regards polarisation within a heterostructure, it should be noted that it is polarisation differences that are relevant [Bernardini and Fiorentini, 2000], as it is such differences that give rise to electric fields: a strained quantum well within a material B will be subject to a total polarisation:

$$\boldsymbol{P} = \boldsymbol{P}^{\mathrm{sp}}_{\mathrm{QW}} - \boldsymbol{P}^{\mathrm{sp}}_{\mathrm{B}} + \boldsymbol{P}^{\mathrm{pz}}_{\mathrm{QW}}(\epsilon), \tag{1.54}$$

This polarisation gives rise to the accumulation of charges at the interfaces between the layers of the structure:

$$\rho_{\mathrm{P}} = -\nabla \cdot \boldsymbol{P} \tag{1.55}$$

where ρ_{P} is the density of charge due to the polarisation, taken into account for the resolution of the drift diffusion model (in Eq. 1.3).

The quantum confined Stark effect (QCSE) [Miller et al., 1984] arises in quantum wells subject to electric fields perpendicular to the layers. As such

fields are inherent in gallium nitride heterostructures due to the charge accumulation at the interfaces between the layers, the quantum Stark effect is critical to their modelling.

Early on in the development of gallium nitride based light emitting diodes it had been observed a band gap narrowing [Nakamura et al., 1995] caused by strain, soon after identified as being due to the quantum Stark effect [Takeuchi et al., 1997].

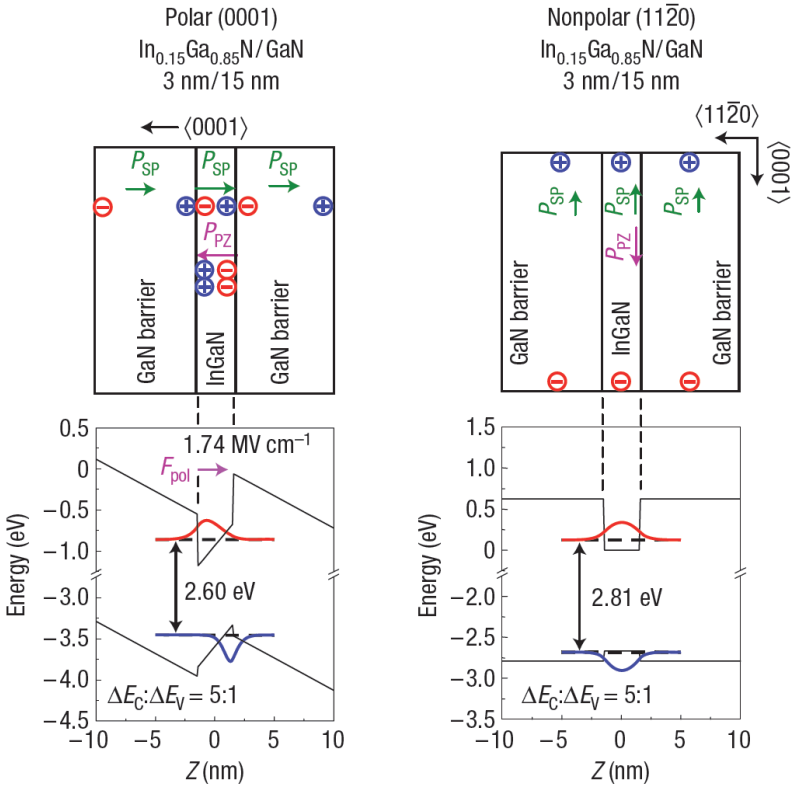


Figure 1.3: From [Chichibu et al., 2006]. Left: Device grown in the polar c direction. Spontaneous and strained-induced polarisation, dominant, induce large electric fields at the interfaces between layers, resulting in a bending of the band structure and the Quantum Confined Stark Effect (see text). Right: Device grown in a non-polar a direction. The polarisation fields are parallel to the plane of the active region, thus not affecting the band structure.

Due to the electric field the band structure becomes tilted (Fig. 1.3) and the energy minima for electron and holes become physically separated on opposite sides of the quantum well. The tilting of the band structure results

in a band structure with triangular wells which energy levels give a smaller gap than their corresponding square wells. This smaller gap is the origin of an energy shift towards lower energies of confined electron-hole pairs. The built-in electric field is also screened when the carrier concentration increases, as the carriers accumulate on opposite sides of the well and generate an electric field that opposes the polarisation-induced field, resulting in a blue shift with excitation power. The separation of carriers results in a reduced wave function overlap, decreasing the probability of both excitonic formation (light absorption) and recombination (light emission) [Seo Im et al., 1998], thereby decreasing the efficiency of the device.

Thus, the mitigation of the quantum Stark effect is central to the optimisation of LEDs, and can possibly be achieved in a number of ways:

- A gradient of aluminium content in the electron blocking layer (EBL) on top of the active region allows to spread the polarisation-induced field [Ryou et al., 2009], that is otherwise sufficiently strong to cause the accumulation of an electron layer between the last barrier of the active region and the EBL, where no holes are present. This quantum stark effect at the EBL interface causes an important concentration of electrons to be unable to interact with holes, thereby impacting the efficiency.

- The use of a quaternary alloy AlGaInN enables the tuning of the polarisation for a fixed band gap, allowing to match the polarisation of adjacent layers [Schubert et al., 2008], reducing the QCSE to the benefit of efficiency.

- Any source of free charged carriers will compensate the built-in electric fields, thus screening the QCSE. The doping of the barriers with silicon has been successfully employed to that effect [Wu et al., 2002], improving the crystal quality at the same time as screening the QCSE. It has been reported however that excessive Si doping can impede the hole transport and thus the structure efficiency [Ryou et al., 2008], requiring a careful optimisation.

- A very active domain of research is the growth of light emitting structures along non-polar or semi-polar directions, that is crystal orientations with a lower polarisation than the standard c (0 0 0 1) orientation. The QSCE can thus be removed [Park, 2002] [Liu et al., 2008a] and thin wells are no longer required, avoiding carrier injections issues that come with it. However, due to the difficulty of growing structures along those directions, their efficiency remains lower.

Quantum transport

The drift diffusion model (Chap. 1.3.1) treats the carriers as fluid-like, only allowing them to move in a local fashion. However, when a feature of the device becomes smaller than the carrier mean free path, it allows them to behave in a non-local fashion and bypass such a small-scaled structure. This phenomenon is particularly important to account for in GaN-based LEDs as the quantum well width is made very narrow to mitigate the quantum confined Stark effect, usually down to ca. 3 nm, which is smaller than the typical electron mean free path of about 5 nm [Stanton et al., 2001]. In the drift diffusion model, carriers flow in a local manner, according to their Fermi level, and carriers able to non-locally escape from the quantum wells do so by violating the normal Fermi statistics, according to the phenomenon of quantum non-locality. This can be interpreted in terms of the hot carrier theory whereby the current flow has a non-local term accounting for some carriers "flying" over the quantum wells [Azoff, 1989]:

$$\boldsymbol{J}_n = \boldsymbol{J}_{\text{loc}}(T) + \beta \boldsymbol{J}_{\text{nloc}}(T), \tag{1.56}$$

where β is a scaling factor.

1.3.4 Numerical considerations

The solving on a computer of the previous equations and the drift-diffusion model (Chap. 1.3.1) in particular requires their discretisation on a mesh. The electrostatic potential of the device obeys Poisson's equation (Eq. 1.3), and the carrier concentrations obey the continuity equations (Eq. 1.12 and 1.13). Semiconductor simulations in general make use of the Scharfetter-Gummel scheme [Scharfetter and Gummel, 1969]. Such finite volume methods exchange the integral of divergence terms over the volume of a mesh node as the integral of their fluxes over the surfaces of the mesh node, using the divergence theorem. Considering a mesh node of volume V and surface δV, the charge inside V due to holes is given by $\int_V qp$ and the rate of flow of electric charge due to holes across δV is given by $\int_{\delta V} \boldsymbol{J}_p \cdot \boldsymbol{n}$, with $\boldsymbol{n}$ the vector normal to δV. The rate of change of charge in V is thus given by [Farrell and Gartland Jr, 1991]:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_V qp = -\int_{\delta V} \boldsymbol{J}_p \cdot \boldsymbol{n} - \int_V qR_p, \tag{1.57}$$

with R the hole recombination rate. Thanks to the divergence theorem the equation becomes:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_V qp = -\int_V \nabla \cdot \boldsymbol{J}_p - \int_V qR_p, \tag{1.58}$$

yielding the following steady-state equation:

$$\int_V \nabla \cdot \boldsymbol{J}_p = -q \int_V R_p \tag{1.59}$$

Instead of solving these 3 partial derivative equations, a $3N$ algebraic system of equations is solved, where the continuous functions underlying the continuous equations are replaced by (discrete) vectors of the value of the function at the mesh nodes. The differential operators are replaced by their discrete counter-parts, difference operators, for instance the continuous equation of the hole current density:

$$\frac{\mathrm{d}J_p}{\mathrm{d}x} = f, \tag{1.60}$$

is discretised in 1D as:

$$\frac{J_{p,i+1/2} - J_{p,i-1/2}}{(h_i + h_{i+1})/2} = f(x_i), \tag{1.61}$$

where x_i denotes the i^{th} node, and $h_i = x_i - x_{i-1}$.

The Scharfetter-Gummel scheme [Scharfetter and Gummel, 1969] then derives the variation of the hole current density as follows:

$$J_{p,i+1/2} - J_{p,i-1/2} = \frac{1}{h_{i+1}} \left(B(\phi_i - \phi_{i+1}) p_{i+1} - B(\phi_{i+1} - \phi_i) p_i \right), \tag{1.62}$$

where $B(z) = \frac{z}{e^z - 1}$ and $B(0) = 1$.

Once the equations modelling the device have been discretised in an algebraic form, the resulting matrix is non-diagonal and the solution must be found in an iterative way using optimisation techniques such as Newton's method, where the zero of a function F is iteratively approximated by:

$$J_F(\boldsymbol{x}_n)(\boldsymbol{x}_{n+1} - \boldsymbol{x}_n) = -F(\boldsymbol{x}_n), \tag{1.63}$$

where $\boldsymbol{x}_{n+1}$ is a better approximation of the zero of F than $\boldsymbol{x}_n$, and J_F is the Jacobian matrix of F: $(J_F)_{ij} = \frac{\delta F_j}{\delta x_i}$.

The considerations above describe how the drift-diffusion model is numerically solved on a mesh describing the device. The numerical simulation of the active region can be achieved in a self-consistent manner, iterating between the resolution of Schrödinger's equation and the drift-diffusion model. A self-consistent carrier density distribution is then achieved, where the density distribution of unconfined carriers is determined by the drift-diffusion

model, and the density distribution of confined carriers is determined by Schrödinger's equation:

$$\begin{aligned} n &= n_{\text{confined}} + n_{\text{unconfined}} \\ p &= p_{\text{confined}} + p_{\text{unconfined}}, \end{aligned} \tag{1.64}$$

where the unconfined densities are determined by Fermi statistics (Eq. 1.10 and 1.11) and the confined densities are determined by:

$$\begin{aligned} n_{\text{confined}}(z) &= \sum_i \phi_n^i(z)\rho_i kTln(1+e^{(E_{fn}-E_i)/kT}) \\ n_{\text{confined}}(z) &= \sum_j \phi_p^j(z)\rho_j kTln(1+e^{(E_{fp}-E_j)/kT}), \end{aligned} \tag{1.65}$$

where i and j run on the confined states as given by solving Schrödinger's equation for quantum wells (Eq. 1.37).

Chapter 2

Machine learning methods

2.1 Supervised learning

Invented as a branch of artificial intelligence in the 1960s and soon after abandoned in favour of expert systems (explicitly programmed computer systems that mimic human experts), machine learning was reborn in the late 1980s with the objective of being an optimisation tool. Early notable achievements have originated from the computer sciences: the recognition of spoken words in 1989 [Waibel et al., 1989], the autonomous driving of a car in 1989 [Pomerleau, 1989], and the playing of backgammon at a competition level in 1992 [Tesauro, 1992] to cite only a few historical landmarks. A recent report presented machine learning as a major driving force for innovation in the years to come [Manyika et al., 2011].

Machine learning algorithms are a kind of artificial intelligence that learn from examples, in an attempt to mimic the way animals learn. Instead of building inferences from past experiences and memories like animals do, the machine learning program builds inferences from a database, that can be seen as formatted "memories" of the machine. To continue on this parallel with animal learning, instead of conditioning the machine with a "treat", the machine is conditioned with a virtual reward when it made the correct inference (or a virtual punishment when it did wrong!). More concretely, given a database the machine learning program tries to find a simple rule or function that explains it best: it checks many different functions and is rewarded when it finds one that can explain the data.

In a broad sense machine learning algorithms deal with the derivation from examples of the relation between an input space and an output space. As the number of examples increases, the reliability of the induction increases, which can be formalised and quantified in a mathematical framework

[Mitchell, 1997]. More practically, it consists in algorithms that automatically learn from data, in the sense of parameters automatically adjusted according to the examples given. Machine learning has proven to be particularly powerful when manually programming a task is difficult or inefficient [Domingos, 2012], which is the case in particular for high-dimensional inputs. Machine learning already is ubiquitous in computer science and business, being heavily used in applications as diverse as web search engines, spam emails filters, stock trading, etc. However, its potential in the physical sciences, and materials science in particular, is only starting to be explored, within a field that is still in its infancy, materials informatics [Rajan, 2005].

A machine learning algorithm, given a training set of examples $(x_i, y_i)_{i=1...N}$ where each x_i is a d-dimensional variable, yields a function g that maps an input x to an output y. The algorithm then has to be tested: the function it yielded must correctly reproduce the examples of a testing set, entirely different from the training set. Each machine learning system can be decomposed into three components [Domingos, 2012]: representation, evaluation, and optimisation. The representation component pertains to both the set of hypotheses $\mathcal{F}$ being explored by the algorithm as possible solutions of the function it is searching as well as how to consider the inputs. The evaluation component pertains to the quantification of the quality of a function $f \in \mathcal{F}$ being considered as a candidate for g. The optimisation component pertains to how $\mathcal{F}$ is explored and consists in the maximisation of the quality of g as determined by the evaluation component.

Although the evaluation and optimisation tools of machine learning algorithms are the same as those of fitting, machine learning has a different purpose: generalisation, which is how well the final hypothesis g performs out of the training sample, optimised using the performance on the training sample as a proxy. This generalisation is ensured by purposely limiting the learning algorithm such that the hypothesis set $\mathcal{F}$ it explores is not too large. Indeed, given all the possible mathematical functions, maximising the evaluation of the final hypothesis will result in a perfect fit of the training data, *i.e.* a memorisation of the training data, with little hope of generalisation on the testing data.

This *overfitting* can be analysed in terms of the bias-variance trade-off [Domingos, 2000]: in the field of machine learning, bias designates the expected generalisation error of the model, and variance designates the tendency of the model to capture noise in the data or "hallucinate" patterns that are absent from the data. A high variance machine learning algorithm would yield very different final hypotheses from similar training data and would have a very small bias, a case which occurs with overfitting. On the contrary a too simple hypothesis set would result in an algorithm yielding

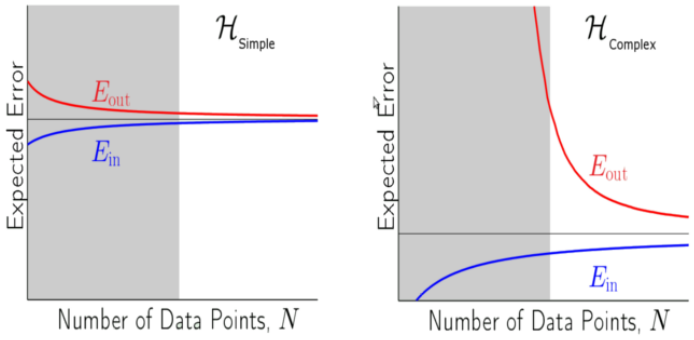


Figure 2.1: From [Abu-Mostafa et al., 2012]. Illustration of the bias-variance tradeoff. Simple models have a higher error for large datasets (high bias) but the error on the training set E_{in} converges towards the error on the testing set E_{out} quickly (low variance). Complex models have a small error for large datasets (small bias), but their error on the training set E_{in} is meaningless for too few data points (high variance), and slowly converges towards the error on the testing set E_{out} for large datasets.

small variance but high bias. The saying "everything should be as simple as it can be, but no simpler" [Calaprice, 2010] is a particularly good guideline in machine learning and can be enforced through quantitative methods known as regularisation [Mitchell, 1997]. The bias variance analysis (see Fig. 2.1) thus offers the surprising possibility to make inductions with quantified guarantees, albeit probabilistic and loose ones: it offers the guarantee that if the hypothesis set contains the true function, then the probability machine learning algorithm returns a poor hypothesis decreases with the size of the training set[Domingos, 2012].

2.2 Active learning

This section is based on the classic book from Sutton and Barto [Sutton and Barto, 1998]. It will be explained more formally in the next section. Active learning (also called reinforcement learning) involves learning from interaction how to behave in order to achieve a goal, by mapping states to actions in order to maximise a numerical reward. The states correspond to situations in which choices have to be made, the actions correspond to the choices selected by the machine, and the rewards correspond to an evaluation of these choices. In contrast to supervised learning, which consists in learning from examples provided to the algorithm, the machine interacts with its

"environment", and must discover by itself which decisions will yield the greatest rewards. The machine is the decision-maker, while the environment corresponds to everything (outside of the machine itself) that interacts with it. The goal of the machine is to maximize the rewards it receives over time.

There are five main components to a reinforcement learning problem: the policy, the reward function, the return function, the evaluation function, and the model of the environment.

• The policy describes the agent's behaviour at a given time: it maps observed states to actions. For instance, given the observed states, is the machine greedy and simply looking for what it thinks is the action that will lead to the best state, or is it allowing for some exploration?

• The reward function associates each possible state (or state-action pair) to a number that measures its desirability. Rewards functions may be used to alter policies; they cannot be modified by the machine.

• The return: the objective of the agent is to maximise its long-term reward. The return is defined as a function of the future reward sequence, and aims at measuring this long-term performance. The goal of the machine is therefore to maximise the return. Commonly, the machine doesn't simply try to maximise the reward of its next action, but tries instead to have an exploratory behaviour such it does not get stuck in local maxima.

• The value function correspond to the expected rewards to be accumulated in the future, i.e. the long-term desirability of states and actions, according to a given policy. They must be estimated iteratively from the sequence of observations. They enable to identify the best possible policies.

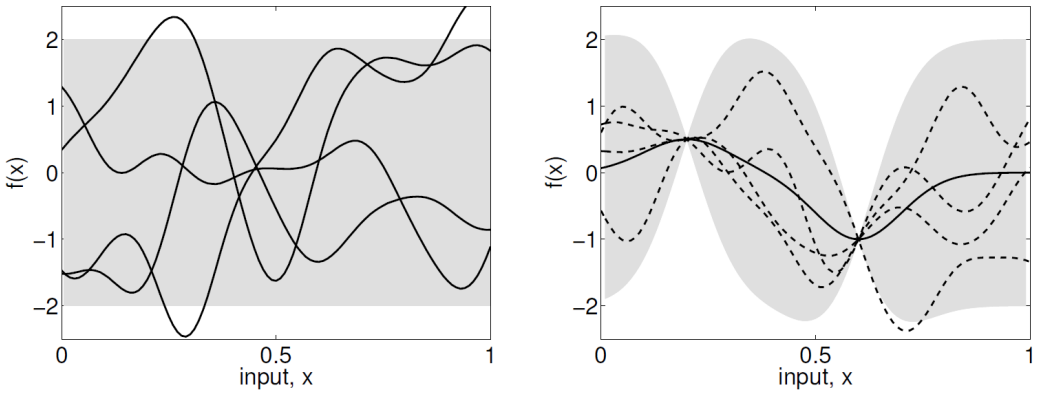
• The model of the environment aims at learning the behaviour of the environment (for instance, being able to predict the next state and reward given the current state). It is useful for planning, i.e. deciding on a path of action by expecting their results before actually implementing them. The accurate model of the environment requires that the machine has an exploratory behaviour.

2.2.1 Exploration versus exploitation

To learn from its environment, the agent must be able to achieve a trade-off between exploration and exploitation. Exploration aims at discovering new actions never used before, while exploitation consists in selecting actions that have yield high rewards in the past. Without exploration, the machine would not be able to identify new actions from which it may benefit from in the future; the balance between exploration and exploitation is therefore crucial, and will be further discussed in chapter 5, 6 and 7.

2.3 Gaussian processes

The use of GPs as a predictive tool started in the 1940's with the basic theory introduced by Kolmogorov and Wiener [Wiener, 1949]. In the geosciences, GP regression is known as kriging [Journel and Huijbregts, 1978]. The work we present in chapter 5 and 7 builds upon a long history of applying GP to emulate computer simulations [O'Hagan and Kingman, 1978, Sacks et al., 1989] and as a component of global optimisation [Journel and Huijbregts, 1978, Gupta and Miescke, 1996, Jones et al., 1998, Osborne et al., 2009, Frazier et al., 2009, Srinivas et al., 2012]. GP regression includes *uncertainties* along with its predictions, making it an ideal model for active learning algorithms.



(a), prior (b), posterior

Figure 2.2: From [Rasmussen, 2006]. Interpretation of the posterior: draw many random functions from the prior, and reject the ones which do not agree with the observations. a) The prior: the space of functions considered as candidates for the model. b) The posterior, the functions selected for their agreement with the data. Where the standard deviation of the resulting model is 0, it indicates a data point.

2.3.1 The function-space view

A GP may be thought of as an infinite-dimensional generalization of a multivariate normal distribution, where each *point* of the black curve of Fig.2.2b is the average of a Gaussian variable, of standard deviation the grey area. GPs for ML fall under the so-called Bayesian modelling framework, where the general knowledge of the function being learned (such as smoothness and variation length-scale) is modelled as a 'prior' distribution over the general

space of functions, and the data serves as a filter over this prior, yielding the 'posterior' distribution of the functions indeed approaching the data (see Fig. 2.2). This posterior distribution over the acceptable functions comes with a mean and a standard deviation for each possible input $\boldsymbol{x}$, that are used respectively as the prediction of the output $f(\boldsymbol{x})$ and the error of this predicted output $\sigma_{\boldsymbol{x}}$.

Gaussian processes are thus giving a distribution over functions, and the inference of Gaussian processes-based machine learning is thus operated by the data, that filters the function space. In this function-space view, a Gaussian process g is defined by its mean m and its covariance k:

$$m(\boldsymbol{x}) = \mathbb{E}[g(\boldsymbol{x})] \tag{2.1}$$
$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(g(\boldsymbol{x}) - m(\boldsymbol{x})) \times (g(\boldsymbol{x}') - m(\boldsymbol{x}'))] \tag{2.2}$$

The value of g at $\boldsymbol{x}$ is thus a random variable of mean $m(\boldsymbol{x})$, and the covariance k gives the distribution over functions of the Gaussian process.

The learning process of Gaussian processes can be seen as randomly drawing functions from the prior distribution, and rejecting all the functions that don't agree with the observed data (see Fig. 2.2).

2.3.2 Implementation

This function-space view is intuitive and helps understanding and interpreting Gaussian processes, but would be very computationally inefficient. The algorithlm we will actually use in the next chapters has been adapted in Python by the Scikit-learn team [Pedregosa et al., 2011] from Lophaven [Lophaven et al., 2002]. In this efficient implementation, a Gaussian process is defined as a constant plus an error term centered around zero:

$$g(\boldsymbol{x}_i) = \mu + \epsilon(\boldsymbol{x}_i) \tag{2.3}$$

where $i = \{1...n\}$ denotes the database points where the function has been evaluated, μ is the average of the process, and $\epsilon(\boldsymbol{x}_i) \sim \mathcal{N}(0, \sigma^2)$ the correlated errors such that $\text{Corr}(\epsilon(\boldsymbol{x}_i), \epsilon(\boldsymbol{x}_j))$ is given by the coefficient (i, j) of the correlation matrix $\boldsymbol{R}$, $R_{i,j}$.

It is this correlation matrix that is the core of this Gaussian process modelling. It can have many different forms, but is generally made from a decaying exponential that makes explicit the diminishing correlation between outputs that have increasingly different inputs:

$$\text{Corr}(\epsilon(\boldsymbol{x}_i), \epsilon(\boldsymbol{x}_j)) = R_{i,j} = \exp(-d(\boldsymbol{x}_i, \boldsymbol{x}_j)) \tag{2.4}$$

with d a distance function, that holds the hyper-parameters of the Gaussian process model.

In this context, the values of μ and σ^2 that maximize the likelihood then have a closed-form expression, which makes the approach computationally efficient:

$$\hat{\mu} = \frac{\mathbf{1}^T \boldsymbol{R}^{-1} \boldsymbol{g}}{\mathbf{1}^T \boldsymbol{R}^{-1} \mathbf{1}} \tag{2.5}$$

$$\hat{\sigma^2} = \frac{(\boldsymbol{g} - \mathbf{1}\hat{\mu})^T \boldsymbol{R}^{-1} (\boldsymbol{g} - \mathbf{1}\hat{\mu})}{n} \tag{2.6}$$

where $\boldsymbol{g} = (g_1, ..., g_n)$ is the vector of the observed outputs. A prediction at a new input $\boldsymbol{x}^*$ is then given by [Sacks et al., 1989]:

$$\hat{g}(\boldsymbol{x}^*) = \hat{\mu} + \boldsymbol{r}^T \boldsymbol{R}^{-1} (\boldsymbol{g} - \mathbf{1}\hat{\mu}) \tag{2.7}$$

with $\boldsymbol{r}$ the correlation vector between the error at $\boldsymbol{x}^*$ and the errors at the already sampled inputs: $r_i(\boldsymbol{x}^*)$ =Corr$(\epsilon(\boldsymbol{x}^*), \epsilon(\boldsymbol{x}_i))$. The first term $\hat{\mu}$ comes from the average of the Gaussian process over all observations, and the second term is the error term that corrects from this average.

2.4 Random forests

2.4.1 Decision trees

Decision trees were invented in the early 1980s simultaneously by Breiman and Quinlan [Breiman, 2001]. Their underlying principle is to build non-linear decision frontiers by using many linear separators, that take the form of hyperplanes. For a set of labelled data $\{(\boldsymbol{x}_n, y_n)\}$, with $n \in \{1, ..., N\}$, with y a label and with $\boldsymbol{x}$ a vector of features with $l = \{1, ..., L\}$ coordinates, these hyperplanes can be written as $\{x_l = c\}$. Figure 2.3 shows an example of such hyperplanes, in the case where $\boldsymbol{x}$ has only 2 coordinates x_1 and x_2. The color of the datapoints represent their label, i.e. the associated value of y; the goal is to learn how to separate datapoints according to their labels. In this example the number of labels is finite (2 possibilities: blue and red), so it's a classification task; in the case where $y \in \mathbb{R}$, it is a regression task. The goal is to learn all the hyperplane separations ("boxes") on the training

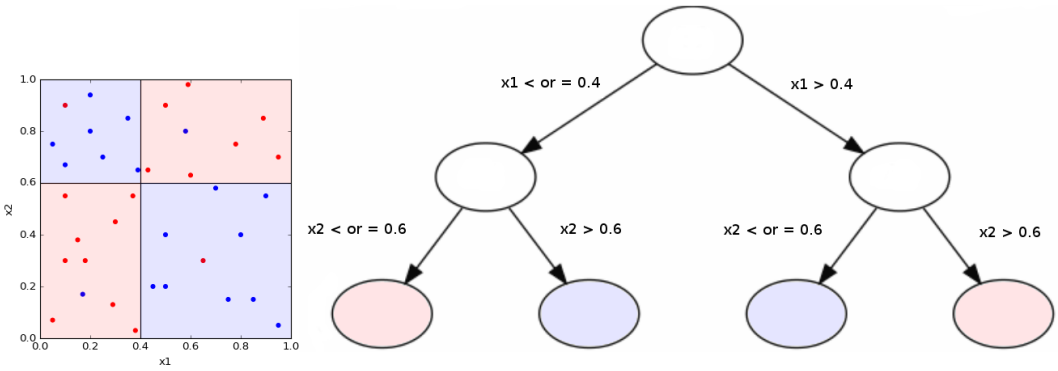


Figure 2.3: Example of splits made by a decision tree. The splits are parallel to the axes, and try to separate as well as possible the blue dots from the red dots.

set; then when a new datapoint (from the testing set) is considered, it will be possible to predict its label according to the area ("box") in which it falls.

The combination of all learnt hyperplanes corresponds to the machine learning function. This constant piece-wise function can easily be represented by a tree, in which each node is associated to a hyperplane (Figure 2.3 also shows the tree corresponding to the illustrated data separation). More specifically, for all $\boldsymbol{x}_n \in \mathbb{R}^L$, this function can be written as:

$$f(\boldsymbol{x}) = \sum_{l=1}^{L} a_l \mathbb{1}_{K_l}(\boldsymbol{x}),$$

where K_l corresponds to a subset of hyperplanes orthogonal to the canonical basis (i.e. one of the "boxes"), and $K_1 \cup ... \cup K_L$ is a partition of $\mathbb{R}^L$. a_l corresponds to the value of the label attributed to the "box" K_l.

• For classification problems, a_l is determined by a majority vote in K_l. In our previous example, if in a given "box" there are more points with a blue label than with a red label, then the associated value of a will be blue. Any testing datapoint falling within this "box" will be given the label blue. Generally speaking, with K the total number of classes,

$$a_l(\boldsymbol{x}) = \text{ arg max}_{k=1,...,K} \sum_{\boldsymbol{x}_n \in K_l(\boldsymbol{x})} \mathbb{1}_{y_n=k}$$

• For regression problems, a_l is determined by the empirical mean of the points in K_l:

$$a_l(\boldsymbol{x}) = \frac{1}{|K_l(\boldsymbol{x})|} \sum_{\boldsymbol{x}_n \in K_l(\boldsymbol{x})} y_n$$

The main challenge to build a tree is to identify where to split the axes in order to build the hyperplanes. Each split is identified recursively, i.e. node after node. At a given node, the procedure aims at identifying the best binary split among a set of $t_{i,\tau}$ possible splits, where i corresponds to the i^{th} coordinate (or axis) of $\boldsymbol{x}$, and τ to the place where to split this axis. There are several ways to choose the values of the τ (for example, they can be drawn according to a histogram, or regularly spaced). The best split within this set will be chosen by optimizing a local loss function L. This loss function is "local" because it is calculated at each node, and its entry parameters vary according to the node. More specifically, at each node and for each split $t_{i,\tau}$, the dataset currently considered S will be split in two subsets (Left and Right), which will be used to evaluate the local loss and select a split. At the next node, the current dataset S will correspond to the Left or Right subset selected at the previous node, which will in turn be split, etc. The procedure ends when a criterion (maximal depth, maximal number of leaves, ...) is reached.

• In the case of classification problems, the choices of the splits involve minimizing the loss function based upon an impurity criterion G. The most commonly used impurity criterion is the Gini index. For a dataset S with K classes:

$$G(S) = \sum_{k=1,\dots,K} \rho_k(S)(1-\rho_k(S)) \text{ , where } \rho_k(S) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{y_n=k}$$

The corresponding loss function is:

$$L(t_{i,\tau}, S) = \frac{N_{\text{left}}}{N} G(\text{Left}(S, i, \tau)) + \frac{N_{\text{right}}}{N} G(\text{Right}(S, i, \tau)),$$

where Left and Right are the two subsets of S separated by the binary split $t_{i,\tau}$, and N_{left} and N_{right} are their respective cardinals. The axis and τ chosen at a given node are $(\hat{i}, \hat{\tau}) = \arg\min_{i,\tau} L(t_{i,\tau}, S)$.

• In the case of regression problems, we maximise a local loss function that aims at measuring the overall variance after the split:

$$L(t_{i,\tau}) = \text{Var}(S) - \frac{N_{\text{left}}}{N}\text{Var}(\text{Left}(S, i, \tau)) - \frac{N_{\text{right}}}{N}\text{Var}(\text{Right}(S, i, \tau)),$$

with the same notations as before, and where Var denotes the empirical variance. Therefore, maximizing this loss function leads to split the data in a way

such that the two subsets created by the split are as homogeneous as possible.

Decision trees have many advantages. In particular, they build a clear decision frontier, which makes the interpretation of results very easy. They are also fast and flexible (e.g. they can be used with continuous and categorial data, for binary/multi-class classification and for regression, ...). They have however a main problem: this estimator has a large variance (in particular, a change at the root implies that the whole tree is different). To alleviate this issue, several machine learning ensemble methods (random forests, extra trees, gradient boosting) consider many trees at the same time. The random forest is one of these methods.

2.4.2 Random forests

Random forests [Breiman, 2001] aim at reducing the variance problem by adding randomness to the construction of the decision trees, and then take an average of the trees' decisions. Two sources of randomness are used while growing the trees:

• Bagging (or Bootstrap-Aggregating): each tree is grown by using only a subset of the original training set, via bootstraping. This means that the data "seen" by each tree is different. It also leads to a greater robustness to outliers.

• Selecting only a subset of axes: while growing a tree, at each node the axes considered for a split are limited to a smaller set of axes (candidate features) F (and no longer to all possible features, as for regular decision trees).

Therefore, B different trees are grown from B (likely) different bootstrap samples, using at each node a randomised subsets of F possible axes (features) for the splits. The output of the random forest is the average over the decisions of the B trees. It is possible to show that the variance of the random forest estimator is smaller than that of a single decision tree. The disadvantages are that this method is slower, and there is a loss in terms of explanatory capability compared to a single decision tree. However, generally speaking random forests remain fast compared to many other machine learning algorithms, and tend to yield very good performances, with little requirements in terms of hyper-parameters optimization.

Chapter 3

Simulation of doping effects in LEDs

3.1 Nano-cathodoluminescence reveals mitigation of the Stark shift in InGaN quantum wells by Si doping

High efficiency InGaN LEDs have vastly better energy efficiency compared to traditional lighting, leading to reductions in energy consumption and air pollutants [Humphreys, 2008]. Current LEDs use multiple InGaN quantum wells (QW) with GaN quantum barriers (QB) between the QWs to confine the carriers and lead to high brightness devices, such as shown in Figure 3.1a. However, a persistent hindrance to the development of higher efficiency devices is the existence of the strong polarisation field across the QWs along the GaN polar axis [0001] [Fiorentini et al., 1999] [Bernardini et al., 1997]. The polarisation induced internal electric fields are responsible for the spatial separation of the electron and hole wave functions in the QWs [Miller et al., 1984] [Ren, 2015], known as the quantum confined Stark effect (QCSE). The QCSE results in a red-shift in the emission and reduced radiative recombination rates, hence making the device less efficient. It has been suggested that the internal electric field can be suppressed by Si doping the QBs [Deguchi et al., 1998] [Cho et al., 1998] [Dalfors et al., 1999], and thus increase the device efficiency. The inclusion of a Si doped layer before the growth of the QWs has also been shown to improve the quantum efficiency [Akasaka et al., 2004] [Otsuji et al., 2006] [Takahashi et al., 2004] [Davies et al., 2015]. Moreover, Kim et al. [Kim et al., 2015] have theoretically shown that achieving the best device efficiency requires optimisation of the Si dopant concentrations

through the QWs.

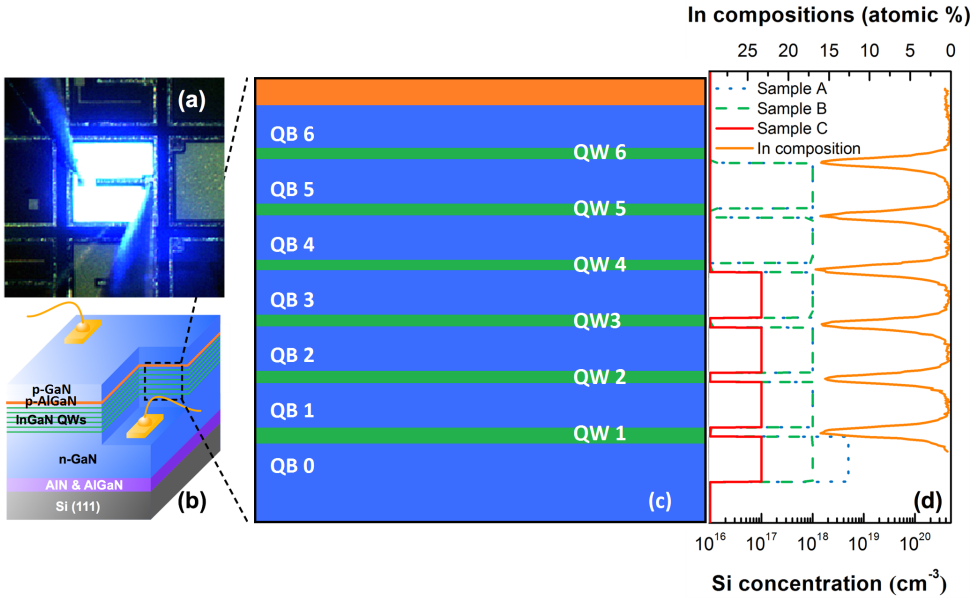


Figure 3.1: Schematic of the LED structures: (a) Photography of the electroluminescence of an LED sample. (b) Schematic of the whole LED structure. (c) Schematic of the active region of the LED. (d) The Si doping profile and In composition across the active region of the samples.

To confirm the simulated properties of the QWs, it is essential to resolve the spectral properties of individual QWs and the influence of dopants on the local emission and electric fields. Nonetheless, common luminescence characterisation techniques, such as electroluminescence (EL) [Park et al., 2007] and photoluminescence (PL) [Deguchi et al., 1998] [Cho et al., 1998], lack the spatial resolution to resolve emission from individual QWs. Cathodoluminescence (CL) in a scanning electron microscope (SEM) has better spatial resolution than EL or PL and has been employed to correlate the luminescence characteristics with structural features. However, the spatial resolution is several tens of nanometers [Petrov and Gvozdover, 1991] [Sonderegger et al., 2006] [Lim et al., 2009] [Merano et al., 2005] and a further improvement in spatial resolution is necessary to study the spectral properties of individual QWs.

Recently, by performing CL in a scanning transmission electron microscope (STEM), a CL spatial resolution of a few nanometers has been achieved by Zagonel et al.; an approach referred to as nano-CL [Zagonel et al., 2010]

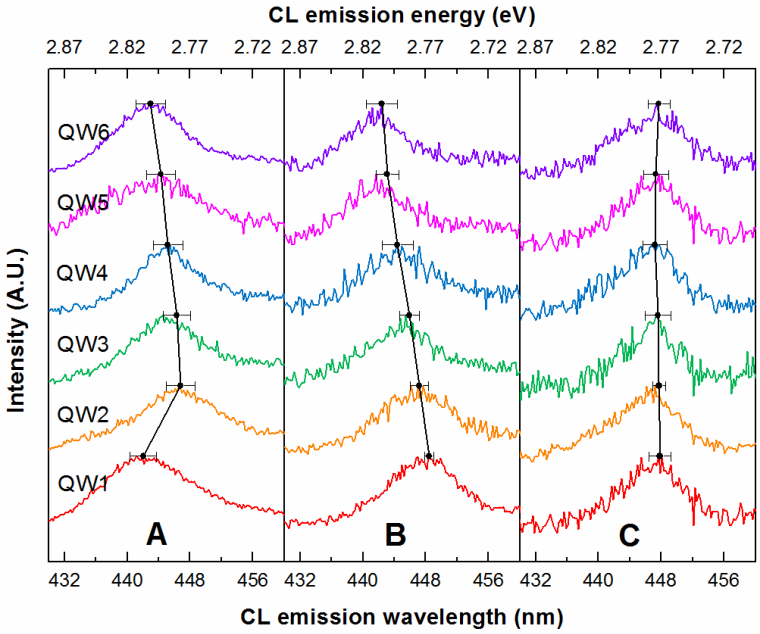


Figure 3.2: Experimental cathodo-luminescence spectra of the samples A, B and C, with mean emission wavelength and standard deviation.

[Zagonel et al., 2012]. Their nano-CL studies on GaN/AlN quantum discs have shown a blue shift caused by quantum confinement for the first time on the same length scale as the device structure. Since then, nano-CL has been used to study spectral properties of GaN quantum dots [Schmidt et al., 2012], stacking faults in GaN nanocolumns [Urban et al., 2015], InGaN insertions in nanowires [Zhou et al., 2015] [Tourbot et al., 2012], and quasi-bulk InGaN [Pantzas et al., 2015]. Nano-CL is thus the only method that can discern the local influence of dopants on the emission properties of individual InGaN QWs.

In this chapter, we use the spectral properties of individual InGaN QWs in LEDs and reveal the influence of doping on the peak emission wavelength [Griffiths et al., 2015]. Nano-CL performed by J.T. Griffiths and S. Zhang is used to resolve the QCSE shifts of individual QWs and the suppression of the internal electric field by Si doping in different LED designs, and Poisson-Schrödinger simulations give us a model of what is happening.

3.2 Simulation of Stark shift mitigation in InGaN quantum wells by Si doping

To understand the correlation between the Si doping and the variation in emission wavelengths, electronic band profile calculations of the experimental structures (Fig. 3.1 and 3.2) are performed using the APSYS simulation package. Materials parameters were taken from the text by J. Piprek [Piprek, 2007]. The band structure was calculated using the 6×6 k $\cdot$ p method [Chuang, 1991] in a finite volume approximation. The carrier transport equations were self-consistently computed and coupled with the Schrödinger equation to determine the confined states in the QWs. Schrödinger and Poisson equations were solved iteratively to account for the band structure deformation with carrier redistribution. We use the literature values for the ionisation energy of Si in GaN (20 meV), as well as a 50% activation fraction [Eiting et al., 1998] [Irokawa et al., 2005] [Hager et al., 2009]. Simulations were performed at zero bias, corresponding to the state examined during nano-CL experiments. Figure 3.3 (right) shows the calculated emission wavelengths of each QW accurately reproduce the variations in the emission wavelength of all three samples within the experimental errors. The calculated electric field at each QW shown in Figure 3.3 (left) shows the same trend observed in the emission wavelength. The mitigation of the QCSE is able to solely explain the observed variation in emission wavelength between each QW and reflects the difference in the local electric field at each QW. We observe that the electric field is uniform across the QWs in sample C with up to 10^{17} cm^{-3} Si doping in the QBs. With 10^{18} cm^{-3} Si doping introduced in samples A and B the electric field decreases towards QW6. An additional reduction in the electric field of QW1 is seen in sample A, with the higher level of doping in QB0. The additional reduction in the electric field of QW1 results in a more even reduction in the electric field across the QWs in contrast to sample B. Among the three structures, sample A shows the smallest overall electric field and strongest mitigation of the QCSE.

To explain the observed variation in emission wavelength and electric field between QWs, the simulated band profiles, charge concentrations, and electric field of sample A are shown in Figure 3.4. The central figure 3.4 shows the polarisation charges are localised at the interfaces between the QW and QB and have the same magnitude across all QWs as they only depend on the QW width and composition. The Si^+ concentration is significantly higher in QB0 due to the increased doping concentration in this barrier. The electron concentration decreases from QW1 to QW6 (Figure 3.4 left), due to the conduction band rising from QW1 to QW6 with respect to the Fermi level

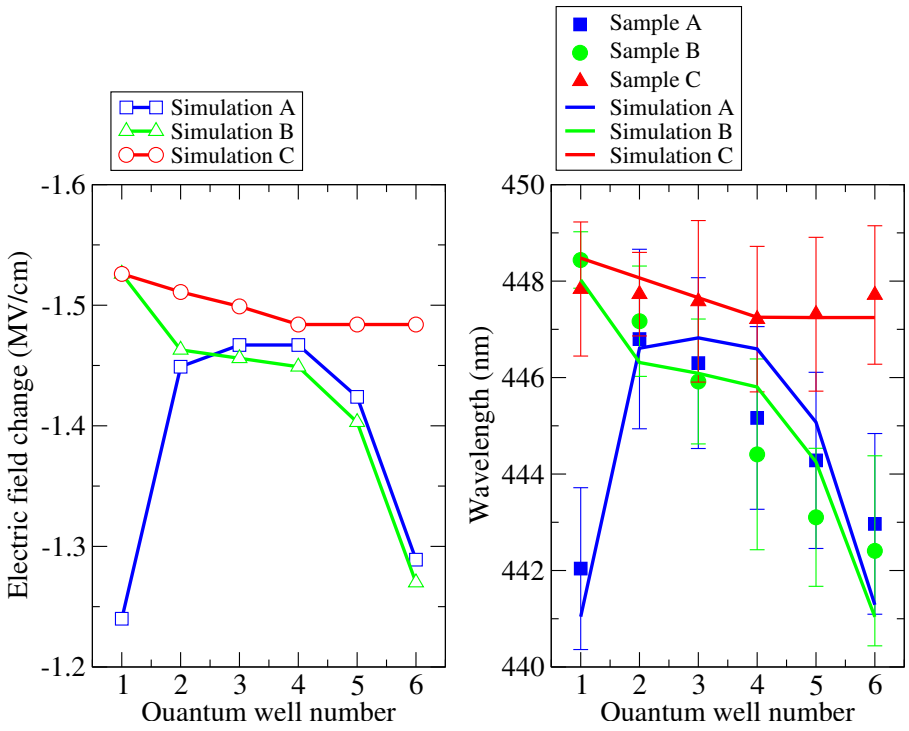


Figure 3.3: Left: The calculated variation in the internal electric field of each QW. Right:The simulated variation in the emission wavelength across each QW, plotted with the experimental data.

(Fig. 3.4 (right)). The hole concentrations are negligible as the Fermi level is significantly higher than the valence band when the structure is unbiased. Figure 3.4 (centre) shows the electric fields caused by the polarisation charges, Si^+, and the free electrons, calculated by the integration of the respective charge concentrations along the growth axis. The electric field caused by polarisation charges is -1.55 MV/cm in each of the six QWs. The Si^+ electric field is of the opposite sign to the polarisation field and hence mitigates the field. However the electric field of the electrons enhances the polarisation field. The significantly larger Si^+ concentration in QB0 of sample A screens the electric field of QW1 and results in the observed blueshift of QW1 with respect to QW2. For sample B, the doping in QB0 is the same as in the subsequent QBs and therefore we do not observe a blueshift of QW1 with respect to the subsequent QWs. The observed reduction in the field over QW2 to QW6 of sample A is due to the reduction in the electron populations. The lower electron concentrations in QW3-6 do not enhance the polarisation field

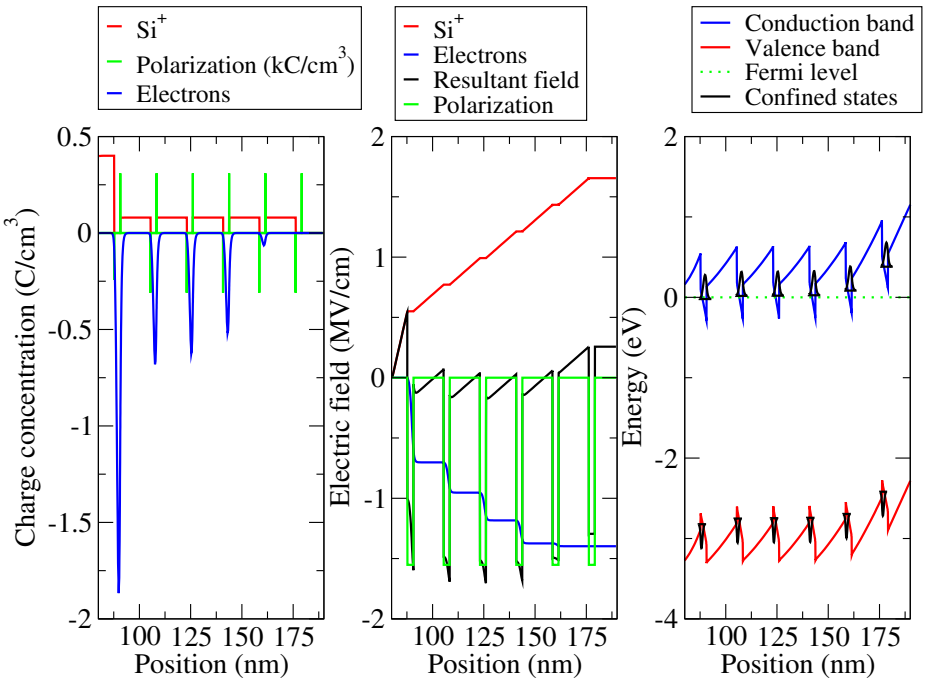


Figure 3.4: For sample A only: Left: The concentration of polarisation charges, electrons, and Si^+ dopants. Middle: The electric field due to the polarisation charges, electrons, Si^+ dopants, and the resultant electric field. Right: The equilibrium band profile with the confined electron and hole states in the QW displayed.

as greatly as QW2, which leads to the observed blueshift in the emission wavelength from QW2 to QW6. Likewise, the blueshift from QW1 to QW6 seen in sample B is due to the reduction in the electron populations from QW1 to QW6, while the Si^+ concentration remains the same for each QB.

3.3 Conclusion

In conclusion, band profile simulations reproduce the experimental variations in the emission wavelengths revealed by nano-CL. Three LED designs are investigated and simulated, that show different mitigation of QCSE in their QWs caused by their different doping profiles. These simulations provide a simple and yet accurate model that attributes the experimentally observed

reduction in the emission wavelength to the mitigation of the QCSE.

As the reduction of QCSE is linked to an increase in efficiency, the association of the analysis of the CL emission of LEDs at the quantum well scale and their accurate modelling provides a valuable tool for the optimisation of light emitting diodes.

The association of nano-CL with Poisson-Schrödinger simulations can thus serve as an approach to study and refine the design of future optoelectronic nanostructures, including the effects from doping and lead to greater improvements in device efficiencies and functionality. This approach can aid in the understanding of the emission properties of a wide range of nanostructures with quantum confinement, in the nitride community and beyond.

Chapter 4

Simulation of inhomogeneous electroluminescence in LEDs

In this chapter, we investigate [Ren et al., 2016] the cause of inhomogeneous electroluminescence (EL) which can be observed in c-plane InGaN/GaN LEDs [Lin et al., 2012b] [Wallace et al., 2014] [Wallace et al., 2015]. We investigate in particular LEDs that exhibit luminescent features of higher brightness than the surrounding material when under bias. The use of Poisson-Schrödinger calculations allowed us to explain the correlation between local emission properties in the LED (experimentally observed by Christopher Ren) and micro and nano-scale structural features, thus offering an explanation to the cause of the inhomogeneous EL.

The LED structures investigated were grown by low-pressure metalorganic vapour-phase epitaxy (MOVPE). The structure studied consists of a low dislocation density (ca. $5 \times 10^8/\mathrm{cm}^2$) GaN template on sapphire with a 2 µm layer of undoped GaN followed by a 3 µm layer of silicon-doped GaN, a 5 period InGaN/GaN multiple quantum well (MQW) region (well width = 3.9 nm) with unintentionally doped GaN barriers (7.6 nm), a magnesium doped AlGaN electron blocking layer (20 nm) and a magnesium-doped GaN cap layer (117 nm). The QWs were grown using the two-temperature - "2T" - method described by Oliver et al. [Oliver et al., 2013], in which reference full details of the growth of the LEDs may be found. Note that the LEDs described in these two papers have a different QW thickness and do not display the same EL inhomogeneity reported here.

EL hyperspectral images were collected by Christopher Ren in order to characterise the local optoelectronic properties at the inhomogeneities. STEM-EDX was performed by Christopher Ren as well on an area of the lamella containing a portion of the V-shaped defect in order to determine any compositional changes the defect may have induced in the LED struc-

ture. Fig. 4.1 b), c), d) show the compositional analysis of the EDX mapping in the area delineated by the red dashed line Fig. 4.1 a). In the gallium (Ga) indium (In) and maps (Figs. 4.1 b) and d)) we observe that the InGaN QWs are significantly disrupted by the hexagonal defect. Additionally, we see in Fig. Figs. 4.1 c) the Al signal overlapping with the QW In signal, and the penetration of In below the QW stack along the pit.

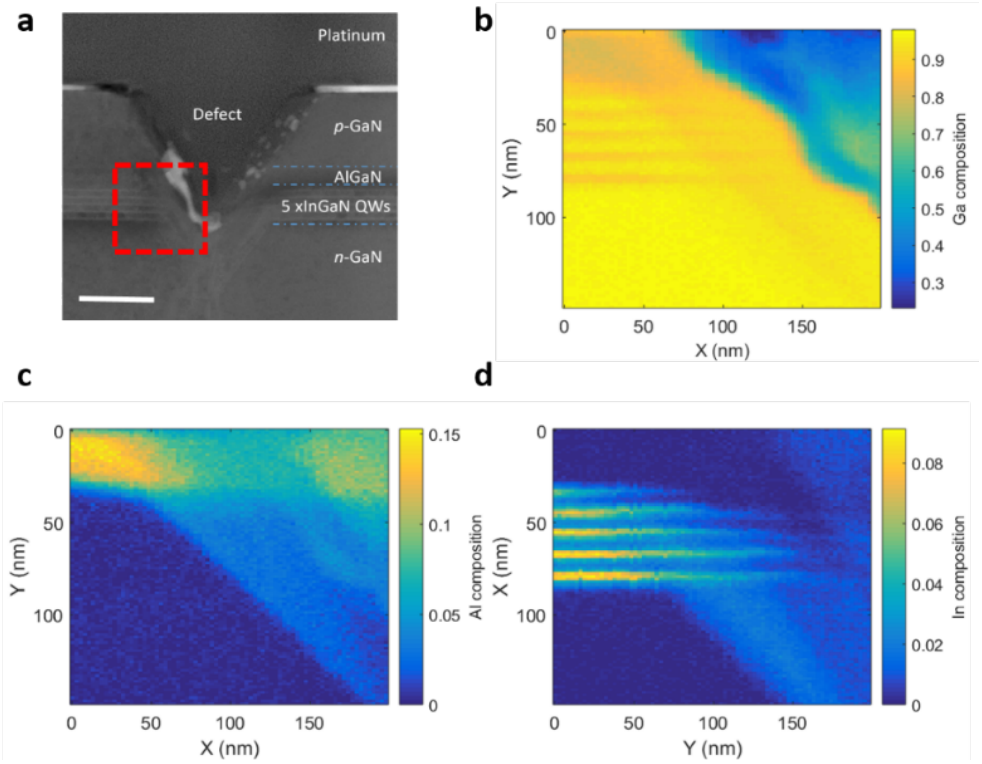


Figure 4.1: a) HAADF-STEM image of a V-shaped defect disrupting the active region of the LED structure, (150 nmscale bar). Cliff-Lorimer analysis of the ternary composition: b) Ga composition c) Al composition and d) In composition. The dashed red square in a) shows the area of b), c), and d).

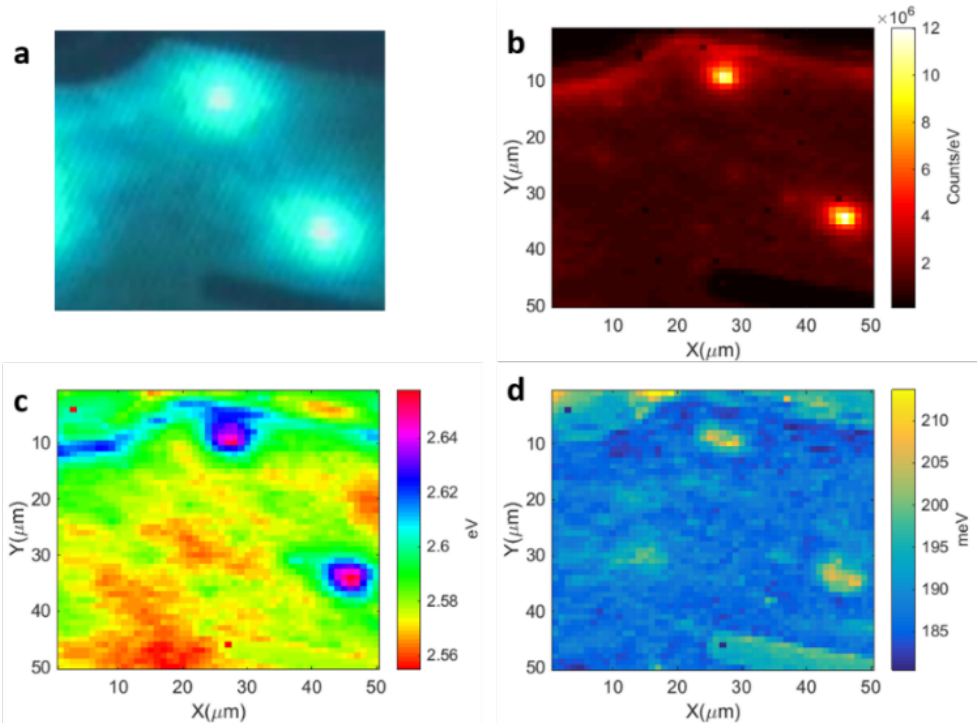


Figure 4.2: a) CCD image of features analysed by EL hyperspectral imaging, b) peak intensity, c) peak energy and d) emission FWHM of the features.

4.1 Simulation of electroluminescence inhomogeneities, as revealed by hyperspectral EL imaging

We investigated the mechanisms by which the presence of the AlGaN laterally adjacent to the InGaN QWs associated with this hexagonal defect could disrupt the EL properties (see Fig. 4.2) of the LED through the use of numerical calculations as described in [Piprek, 2014]. The LED structure was simulated with similar dimensions to our devices, with the presence of the V-shaped defect simulated, as shown in Fig. 4.3.

The results of the calculations are shown below in Fig. 4.4. We find that the presence of the V-shaped defect results in an area with a high radiative recombination rate adjacent to the defect. The mechanism behind this effect can be determined by examining the electron and hole concentrations across the active region in the area neighbouring the V-shaped defect (red lines in

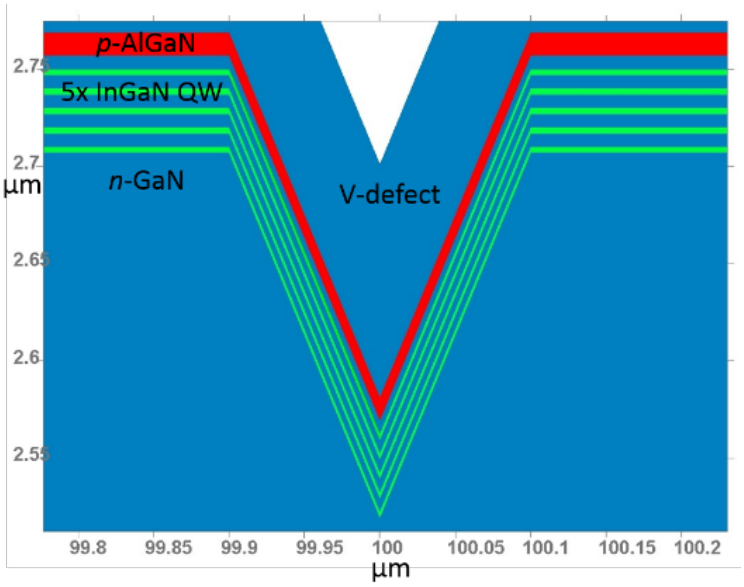


Figure 4.3: Simulated LED structure, with an incorporated V-pit defect disrupting the active region.

Fig 4.4) and further away (black lines in Fig 4.4). The electron concentration (Fig 4.4 c) close to the inclusion remains largely unchanged compared to the surroundings, whereas the hole concentration in the lowest QW (Fig 4.4 (d)) is enhanced, resulting in an increased radiative recombination rate as shown in Fig. 4.4 b).

As such, the numerical calculations suggest the presence of this type of inhomogeneity in the EL can be explained by the increased hole injection in the area around the V-shaped defect. Our model shows that this is due to the influence of what can be seen as an effective lateral heterojunction being formed in the MQW by the presence of AlGaN in the defect. This explanation is supported by the hyperspectral EL data shown in Fig. 4.2: the spectral blue shift observed in Fig. 4.2 (c) and the emission line-width broadening observed in Fig. 4.2 (d) both of which may be due to valence band filling near the V-pit by the excess hole concentration. It is nonetheless important to note that the contribution of the bundle of TDs to local conductivity [Han et al., 2013] [Shan et al., 2011] has not been incorporated in this model. We also note that the lateral dimension and relative brightness of the feature depend on the p-doping of the AlGaN sidewalls of the defect, however the feature exists across a range of dopant concentrations examined in our model.

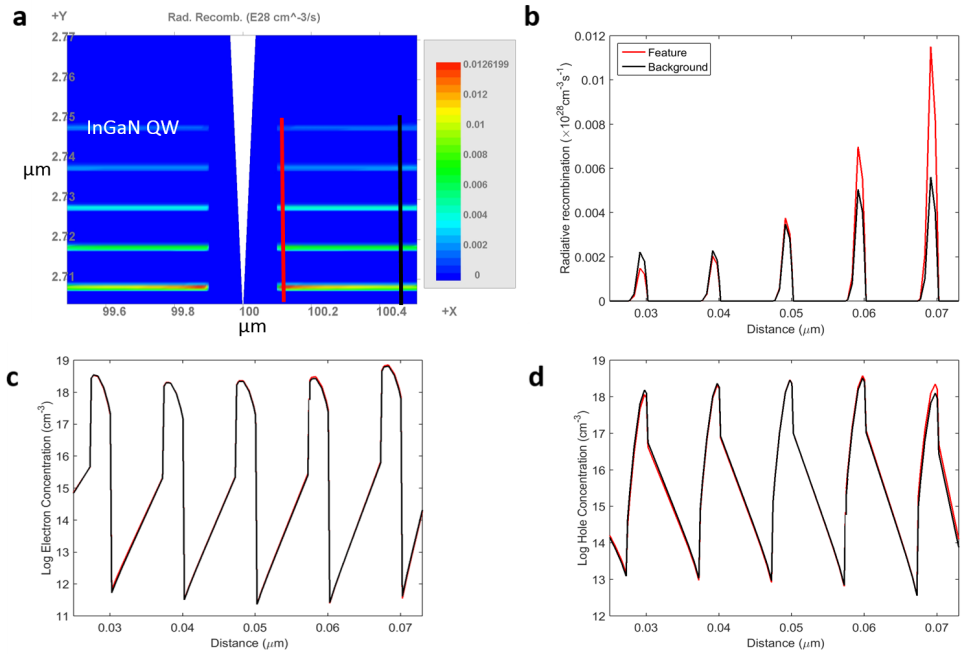


Figure 4.4: APSYS calculations results a) Radiative recombination events b) radiative recombination profiles c) electron concentration d) hole concentration. Red and black traces correspond to the profiles shown in a). In b), c) and d), the p-side top well is to the left.

4.2 Conclusion

In conclusion, Poisson-Schrödinger calculations suggest that the mechanism behind the bright features can be explained by the enhanced injection of holes in the active region due to the presence of AlGaN in the V-shaped defect.

Our analysis suggests that careful control of GaN template growth is key to avoiding such lateral inhomogeneity in the EL from these devices.

The association of multi-microscopy analysis with Poisson-Schrödinger simulations show that disruptions of the active region with structural defects can lead to a detrimental accumulation of carriers. We believe this approach can generalise to the analysis and simulation of a wide array of structures where the quantum confinement is disturbed.

Chapter 5

Parametrisation of simulations from experimental results using machine learning

5.1 Background on simulation design by active learning

Reinforcement learning for calibrating simulations is an emerging field, and the number of papers exploring this idea is relatively small. The methods proposed usually consist of a supervised predictive modeling coupled with an active learning strategy to select the next datapoints (the next simulations to run); each new simulation adds a new datapoint to the training data of the predictive model. The predictive algoritm's input typically corresponds to the parameters of the simulation, and its output to the result of the simulation.

Gramacy [Gramacy et al., 2004] uses a hierarchical predictive model: the input data are partitioned into sub-samples by a decision tree, and a Gaussian process model is built on each of the sub-samples. Then the following data point (next simulation to be run) corresponds to the one with the largest standard deviation in terms of predicted output. In a similar way, [Pfingsten, 2006] uses Gaussian processes as a predictive model, and the next simulations to be run are also selected according to the predictive model's variance. The author tests the algorithm on three different finite elements simulations of electro-mechanical devices, and shows that for a given number of iterations a much higher accuracy can be reached with this strategy as compared to passive learning.

Burl [Burl and Wang, 2009] applies several different approaches to a toy simulation consisting in balancing the stick of an inverted pendulum in order to reach a vertical position. At each iteration, a supervised learning algorithm is built to create a predictive model from the currently available training data; its inputs are the simulations' parameters, and the output corresponds to wheather the simulations were successful (+1) or not (-1). Three different types of predictive models are tested: support vector machines, kernel density estimation, and Gaussian processes. The authors also compare several active learning strategies to pick the next simulation to be run: passive learning (random choice), most confused point (choice of the most uncertain point given the current state of knowledge), most informative point (choice based on the largest information gain), and a meta-strategy (random choice given a weighted preference over the points).

Naik [Naik et al., 2013] [Naik et al., 2016] applies an active learning strategy on both simulation and experimental data of chemical compounds for drug development. Their approach relies on clustering the simulation / experimental results, and using a correlation analysis to select which simulations / experiments to run next. They show that they are able to identify the effect of chemical compounds while performing only a fraction of all possible simulations / experiments. Cevik [Cevik et al., 2016] uses reinforcement learning to calibrate cancer simulations. They use an ensemble of neural networks as predictive model. This model takes as input the simulations' parameters, and as output a score measuring the conformity of a simulation's results to observed outcomes. The neural network model is then used to identify a batch of points with the best predicted scores, and the corresponding simulations are run and added as new training points. By using this strategy, the authors show that they require 100 times fewer trials to identify the 69 parameter sets found to match observed data in a previous study, as compared to a brute-force approach.

5.2 Active machine learning coupled with simulations

We use an approach similar to our work in [Rouet-Leduc et al., 2016]. This approach has two main components: the use of a machine learning regression to predict the output of the simulation (given its parameters), and the EGO

[Jones et al., 1998] heuristic to select the next sample point (i.e. the parameters of the next simulation to run). EGO (see chapters 6 and 7 for more details) aims at minimizing the expected distance between the simulation and the experiment, accounting for the machine learning model's uncertainty. Let D_n be the dataset constituted of $\boldsymbol{x}_n$ the parameters of all simulations run so far, and y_n the associated distances. With y_{max} the best distance in the dataset D_n, the expected improvement in distance for a new y is expressed as:

$$\begin{aligned} v(\boldsymbol{x}, \mathcal{D}_n) &= \int_{y_{\max}}^{\infty} (y - y_{\max}) P(y|\boldsymbol{x}, \mathcal{D}_n) \mathrm{d}y \\ &= \frac{\sigma}{\sqrt{2\pi}} \left(e^{-\alpha^2} + \sqrt{\pi}\alpha \operatorname{erfc}(-\alpha) \right) \end{aligned} \tag{5.1}$$

where $\operatorname{erfc}(\cdot)$ denotes the complementary error function, μ and σ denote the empirical mean and standard deviation of the machine learning model, and $\alpha = (\mu - y_{\max})/\sqrt{2}\sigma$ is the scaled difference between the expected distance of $\boldsymbol{x}$ and the highest distance in our dataset. In summary, EGO aims at selecting the predicted best simulation while taking into account model uncertainty (exploring regions of space where the algorithm is unsure, to avoid being stuck in local minima).

Then the parameters chosen for the next simulation to be run are:

$$\boldsymbol{x}_{n+1} = \arg\max_{\boldsymbol{x}} v(\boldsymbol{x}, \mathcal{D}_n) \tag{5.2}$$

The EGO algorithm was initially developed in association with Gaussian processes, and Gaussian processes were used in [Rouet-Leduc et al., 2016]. We implemented the algorithm with Gaussian processes and random forests (RF). This second choice is motivated by two main arguments: since an RF model is an ensemble model built out of many decision trees, it is straightforward to retrieve an empirical mean and standard deviation from this ensemble of models. This is what we use for μ and σ. Furthermore, Scornet recently showed [Scornet, 2016] that under some assumptions, random forests converge towards Gaussian processes when the number of decision trees goes to infinity, which also justifies the use of Eq. 5.1 here.

5.3 Fitting a model to experimental data with reinforcement learning

To develop the method we used the APSYS Poisson-Schrödinger simulation package to model experimental results obtained by M.A. Hopkins from the University of Bath, from a 10 quantum wells GaN/InGaN LEDs grown by M. Kappers from the University of Cambridge.

The critical part of this automatised parameter tuning algorithm is to identify a suitable metric to measure how close the modelled and experimental data are.

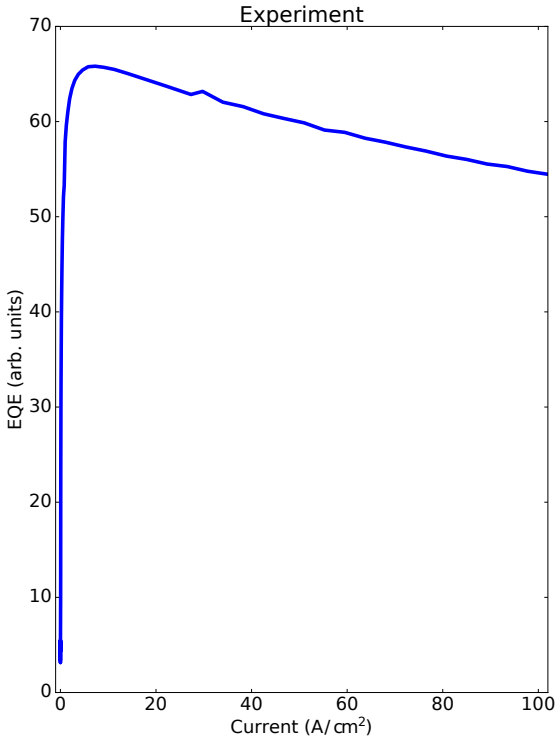


Figure 5.1: Experimental External Quantum Efficiency (EQE) from a 10 quantum wells GaN/InGaN LED.

We build our own measure of "resemblance". The experimental data we use here only give us external quantum efficiency (EQE) in arbitrary units as a function of current density (Fig. 5.1). For the sake of demonstrating the fitting A.I. developed in this chapter, we arbitrarily assume their peak

IQE to be around 65%, and neglect variations of extraction efficiency with current density.

We then use as a measure of distance d between simulation and experiment the average distance in IQE between the two at a finite set of current densities $\{I_i\}_{i=1...n}$:

$$d = \frac{1}{n} \sum_{i=1}^{n} |\eta_{\mathrm{IQE,exp}}(I_i) - \eta_{\mathrm{IQE,sim}}(I_i)| \tag{5.3}$$

This measure serves as the label y of the machine learning regression.

We note here that adapting this approach to new experimental setups would require tuning the metric used here, or building a new one altogether.

In line with our work in [Rouet-Leduc et al., 2016], the reinforcement algorithm has two main steps:

First, a number of simulations are run, with parameters selected via a latin hypercube. Latin hypercube sampling aims at generating pseudo-random parameter values from a multidimensional distribution such that the multi-dimentional volume is uniformly sampled. The machine learning regressor uses these first points to build an initial model, linking parameters of the simulation to the corresponding distance (label).

Then the iterative process begins. The best parameters to explore next are selected through the EGO heuristic (see chapter 6 and 7 for details), and the corresponding simulation is run. The associated distance is then added to the database, and the machine learning model is updated. This step continues until we consider that the model is strong enough.

Besides minimizing the distance, this approach allows to build a model that can predict the performance of the simulation (distance from experiment given by Eq. 5.3) according to its parameters. As the model becomes good enough, the advantages in terms of computation cost become obvious, as predicting simulation outputs with Gaussian processes or random forests is several orders of magnitude faster than running a full-scale Poisson-Schrödinger simulation. To evaluate this model, we draw 20 points with parameters selected uniformly at random, that we use as a test set. In the following sections, we measure the score of the model through the R^2 score. We optimized hyper-parameters by 5-fold cross validation on the hypercube data, and kept these optimized parameters afterwards.

Fitting the model to the experiment

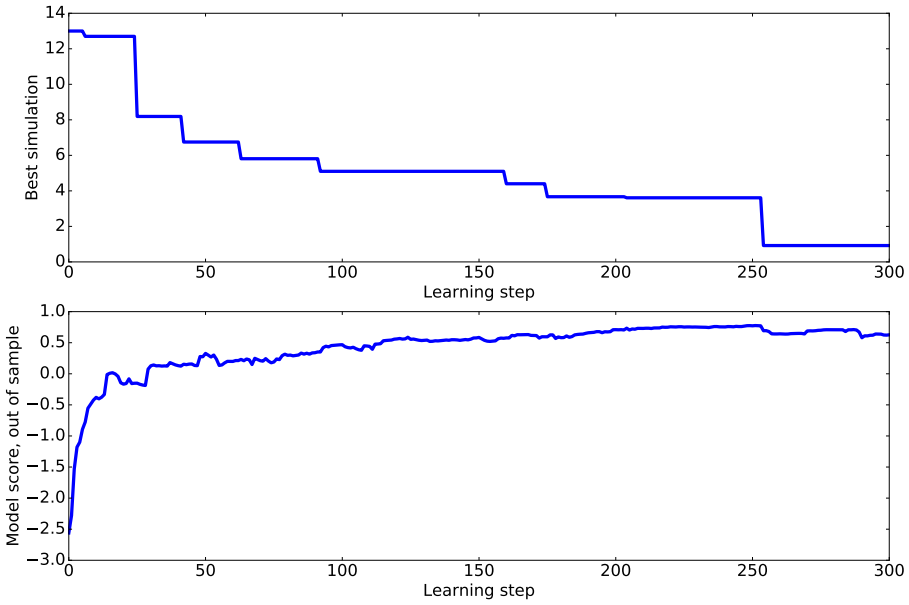


Figure 5.2: Automatic fitting in a 6-dimensional space using Gaussian processes. The input of the GP model consists in 6 different parameters of the simulation, that are allowed to vary on a range that spans 2 orders of magnitude. The output is the distance between the resulting simulation and the experiment, according to according to Eq. 5.3. The database is built by running the Poisson-Schrödinger simulations, and the GP model becomes a proxy for those simulations, that gets more and more accurate as the active learning progresses. Upper figure: simulation closest to the experiment so far. Lower figure: R^2 score of the GP model that predicts how a new simulation is going to be close to the experiment, as a function of the active learning steps. The score is evaluated on a testing set of 20 simulations that have random parameters.

The use of Gaussian processes and the active learning algorithm described above enables us to automatically fit the model to the experimental results from M.A. Hopkins and M. Kappers. The aim here is to find the parameters

of the model that will lead to similar efficiency curves.

In this example, the algorithm relentlessly explores a 6-dimentional parameter space that includes: the polarisation screening, the Auger coefficient of non-radiative recombination, the lifetime coefficient of the carriers, the scattering coefficient of the carriers in the active region, and the exchange coefficient that normalizes the bandgap to account for exchange effects. In this high-dimensional space it is able to automatically find physical parameters that very closely match the experimental results, as seen on Fig. 5.2.

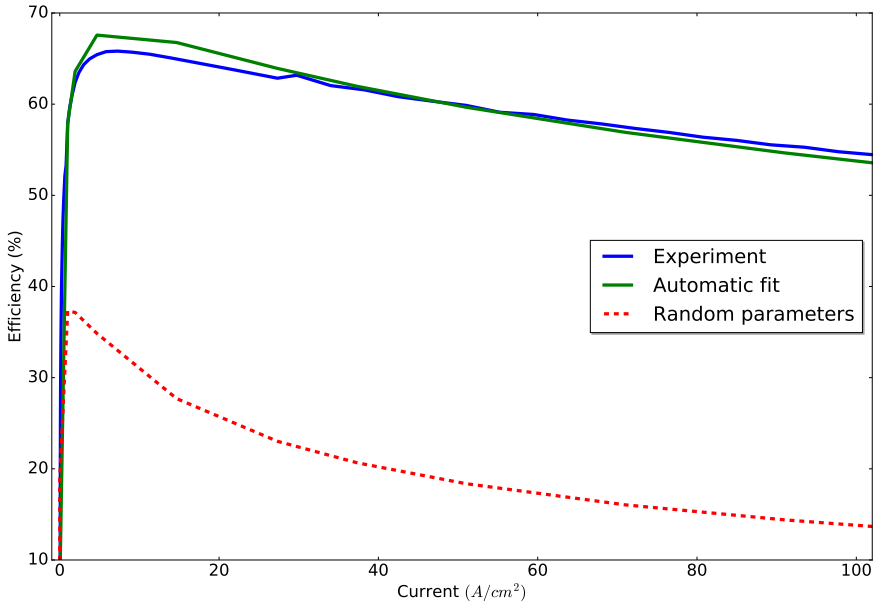


Figure 5.3: Result of the machine learning-based parametrisation of the Poisson-Schrödinger simulation. A simulation using random parameters is shown to give a sense of the variability of the simulation outputs.

We compared the performance of Gaussian processes and random forests. Gaussian processes are much faster at finding small distances, and tend to build a better model than random forests (their R^2 is higher). Generally speaking, Gaussian processes can be very efficient when there are few datapoints, and the true underlying model is rather continuous.

Interestingly the R^2 curve as a function of the active learning step of the Gaussian processes model shows drops. These drops are not nearly as marked with random forest models.

This is likely to be because there is a change in input distribution when shifting from the latin hypercube data to the EGO data, and prior distribution is crucial with Gaussian processes.

In the end, the best modelled stress curve looks very similar to the experimental data, and is extremely close according to our metric.

On Fig. 5.3 we show the experimental results we are fitting, the simulation results of the fitting algorithm, and a simulation ran with random parameters to give an idea of the variability of the simulations.

5.4 Conclusion

In this work we showed machine learning is a very powerful tool for setting up simulation parameters to agree with experimental results. This setting up can be done automatically, removing a labour intensive aspect of simulations. The algorithm intelligently navigates the parameter space to minimise the overall number of simulations required to fit experimental results, by interpolating missing simulations from existing ones.

This automatic tuning of the simulations to experiments enables to easily integrate new experimental results, thus making the simulations more robust and quantitatively accurate.

In addition to automatically and efficiently prepare simulation parameters, the method developed here produces a model that maps physical parameters to the ability of the simulator to reproduce the experiment. This enables the visualisation and a better understanding of the effects of the parameters in the simulation model.

Chapter 6

Convergence of opto-electronic devices simulations using machine learning

In order to run Poisson-Schrödinger simulations of optoelectronic devices (such as LEDs or transistors), one must solve equations of both classical and quantum length scales. These simulations often include many parameters; according to their values and the device simulated, convergence can be difficult and may require a laborious process involving trial and error.

Here we try to accelerate the Poisson-Schrödinger convergence process, by relying on machine learning and statistical methods to suggest new parameters to be sampled in a highly targeted way. To this end, we use the Efficient Global Optimisation (EGO) strategy of Jones *et al.* [Jones et al., 1998]. The EGO strategy allows us to iteratively select sample points in order to maximise the expected improvement in convergence, while simultaneously accounting for model uncertainty. Our approach relies on a "database" of simulations with known numerical and/or physical parameters (i.e. mesh density and geometry, solver parameters, etc.), and labelled by the resulting progress towards convergence. Using this database, we build a ML regression model to predict the convergence of unseen simulations. The algorithm proceeds iteratively: after a new simulation is run, its result is added to the database and the ML model is updated. A greedy approach to experimental design would consist in sampling only the simulations for which the ML model predicts maximum convergence. Here however, we need to reach a more delicate balance between exploration and exploitation, in order to find global optimum parameters. Therefore we also explore parameters for which the ML model is uncertain. This approach improves the global accuracy of the ML model, and avoids being trapped in a region of the parameter space

for which the convergence is only locally maximal.

6.1 Random forests to model convergence

Our approach successively relies on a Random Forest (RF) regression to predict convergence rates, and on the EGO heuristic to select the next sample point, that maximises the expected improvement in convergence rates according to the RF model uncertainty.

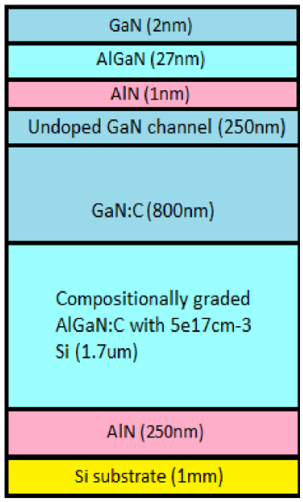


Figure 6.1: Structure of a GaN transistor device, simulated with the help of the automated converger.

Random Forests [Kleinberg, 1996, Ho, 1998, Breiman, 2001] are a popular machine learning algorithm for both regression and classification. Applications of random forests, or machine learning in general, to parameter optimisation in algorithms is an emerging field [Hutter et al., 2011, Bergstra et al., 2011, Rouet-Leduc et al., 2016], and to our knowledge this is the first attempt to apply it to the convergence of optoelectronic simulations.

Here, the predictive output of our regression model corresponds to the fraction of convergence for a given simulation: 0% means that the simulation did not converge at all, while 100% corresponds to a simulation that converged successfully. This percentage is determined by the completion fraction of optoelectronic simulations, where voltages or currents are increased up to target values. For example, a simulation where the drain to source voltage was to be ramped up from 0 to 20V, and that failed at 15V, would

have converged at 75%. This method readily extends to simulations that have a target error instead.

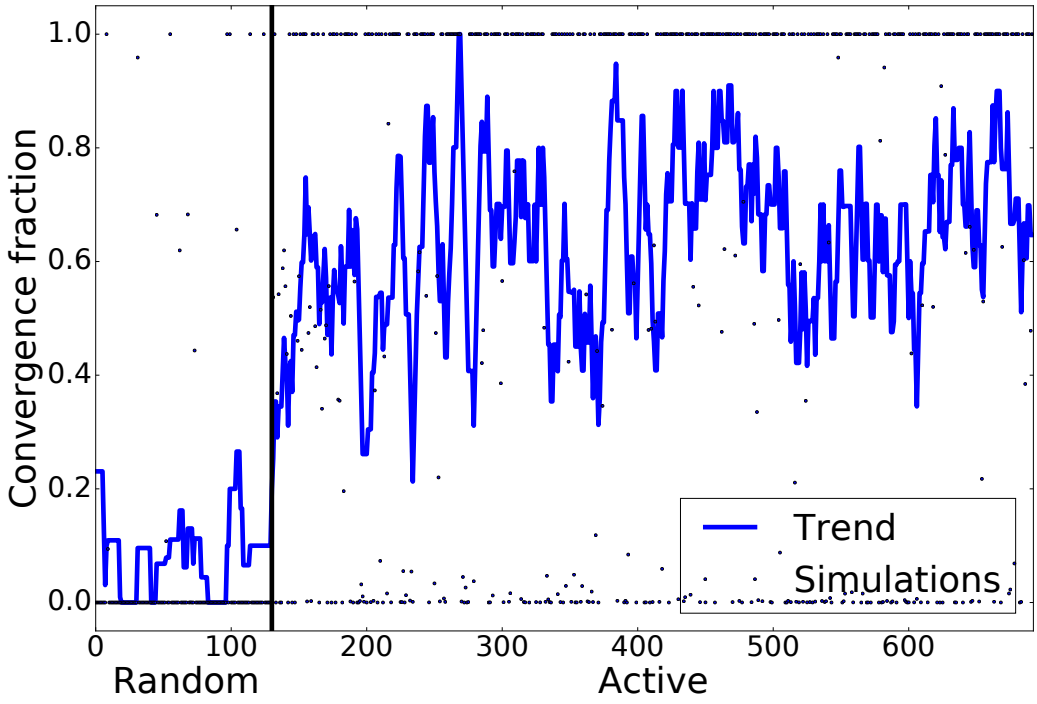


Figure 6.2: Example of application of the active learning for the convergence of simulations of GaN transistors. With random numerical parameters, $\approx$ 15% of the simulations converge. Once the active learning begins (black vertical bar), within a few iterations most of the simulations fully converge.

A RF corresponds to an ensemble of many decision trees; the prediction of the forest is an average of the predictions of each individual tree. Since they are built from many decision trees, RFs allow to quantify uncertainty. Here we work in the space of numerical and physical parameters $\boldsymbol{x}$, and the RF is used to model the probability distribution of the convergence ratio $y = f(\boldsymbol{x})$.

We use the `RandomForestRegressor` module of the scikit-learn Python package [Pedregosa et al., 2011].

The RF model maps a dataset of numerical and physical parameters $\mathcal{D}_n = \{(\boldsymbol{x}_i, y_i)_{i=1...n}\}$ to its associated convergence ratios. More specifically,

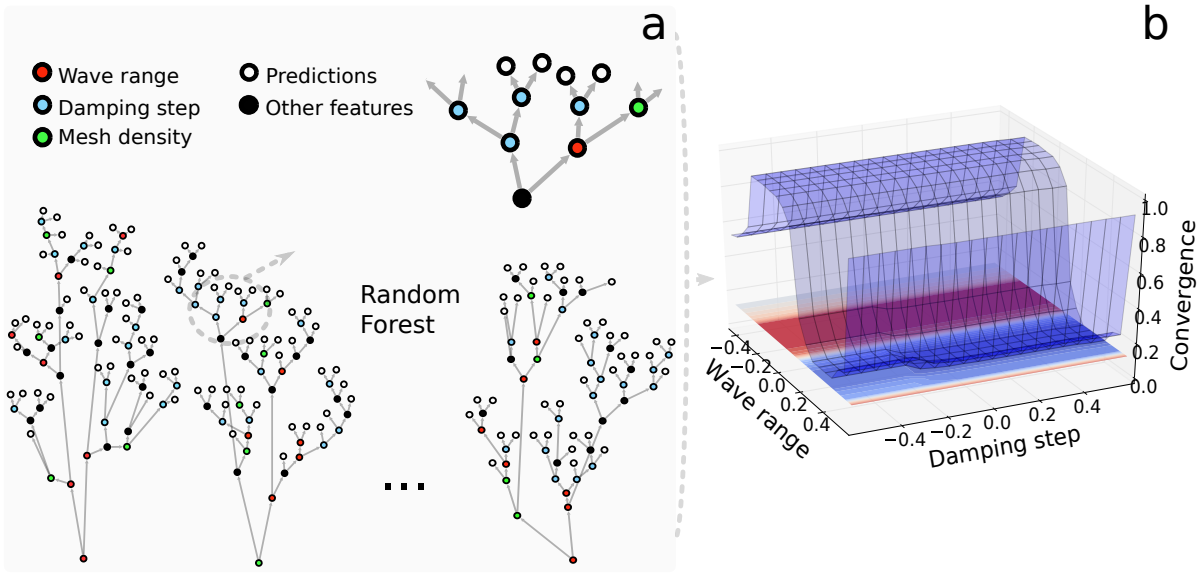


Figure 6.3: **a**: Representation of the random forest that constructs our predictive model. Each tree node indicates a split, and the color indicates which feature is used to make the decision. **b**: The predictive model as a function of 2 variables.

each RF model is built from a dataset $\mathcal{D}_n = \{(\boldsymbol{x}_i, y_i)_{i=1...n}\}$ containing simulation parameters $\boldsymbol{x}_i$ and the corresponding convergence ratios y_i. Once trained, this model can predict the convergence ratio y for a new set of parameters x. Then the question of future simulations arises: what is the best approach to select new parameters $(\boldsymbol{x}_{n+1}, \boldsymbol{x}_{n+2}, ...)$ such that the associated simulations achieve better convergence? Here the EGO strategy comes into play: at each step in the algorithm loop, the next selected parameters should optimise the *expected* improvement in convergence, accounting for model uncertainty [Jones et al., 1998, Lophaven et al., 2002, Rouet-Leduc et al., 2016]. With $y_{\max}$ the best convergence ratio currently in the dataset $\mathcal{D}_n$, this expected improvement in convergence may be expressed as:

$$\begin{aligned} v(\boldsymbol{x}, \mathcal{D}_n) &= \int_{y_{\max}}^{\infty} (y - y_{\max}) P(y|\boldsymbol{x}, \mathcal{D}_n) \mathrm{d}y \\ &= \frac{\sigma}{\sqrt{2\pi}} \left(e^{-\alpha^2} + \sqrt{\pi} \alpha \, \mathrm{erfc}(-\alpha) \right) \end{aligned} \tag{6.1}$$

where $\mathrm{erfc}(\cdot)$ denotes the complementary error function, $\alpha = (\mu - y_{\max})/\sqrt{2}\sigma$ the scaled difference between the expected convergence ratio of $\boldsymbol{x}$ and the

highest convergence ratio in our dataset, and σ the empirical standard deviation of the RF model (see Chap. 5).

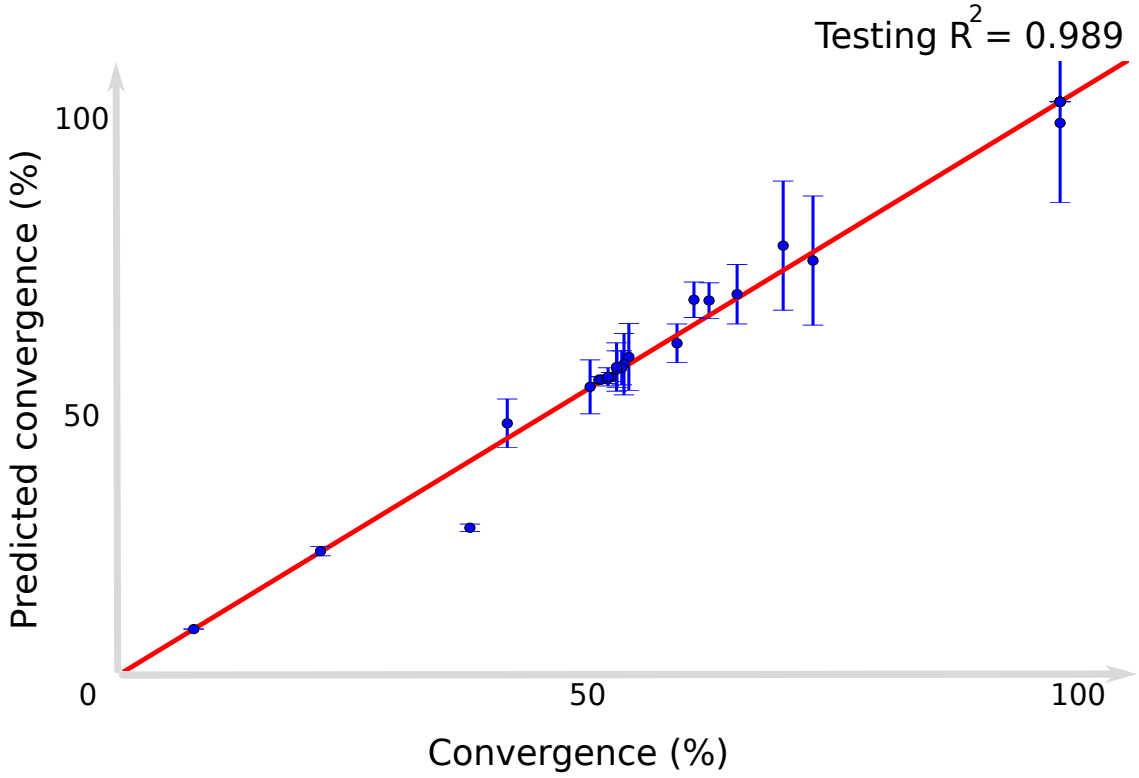


Figure 6.4: The active learning algorithm builds a very accurate model mapping the numerical parameters to the convergence. The above results show the predicted convergence for out of sample simulations (data never seen by the algorithm). The input of the model consists in 5 parameters of the simulation, and the output is the predicted fraction of convergence. The simulation parameters are allowed to vary within a range that spans 3 orders of magnitude. The database is built by running the Poisson-Schrödinger simulations. The R^2 is evaluated on a testing set of 20 simulations that have random parameters. The error bars show the uncertainty of the prediction, as given by the empirical standard deviation of the predictions from the decision trees that constitute the random forest. The uncertainty of the model is higher for the simulations that have a high convergence rate, as they are rarer than the failed simulations in the training set.

We select the next sample point to optimise the objective function v:

$$\boldsymbol{x}_{n+1} = \arg\max_{\boldsymbol{x}} v(\boldsymbol{x}, \mathcal{D}_n). \tag{6.2}$$

The selected numerical parameters $\boldsymbol{x}_{n+1}$ are used as input in a Poisson-Schrödinger code and its associated convergence ratio y_{n+1} is returned by the APSYS software. The next step consists in extending our dataset $\mathcal{D}_{n+1} = \mathcal{D}_n \cup \{(\boldsymbol{x}_{n+1}, y_{n+1})\}$, and updating the RF model $P(y|\boldsymbol{x}, \mathcal{D}_{n+1})$ accordingly. The updated model is used to select *another* sample point via Eq. (6.2). This iterative process continues until convergence can be reliably achieved. It allows to simultaneously build a predictive ML model of simulation convergence, over a broad range of numerical and/or physical parameters.

When the model reaches a high degree of certainty, the objective function primarily seeks to select new points $\boldsymbol{x}$ with higher expected convergence μ than the best known $y_{\max}$. In contrast, when there is no obvious opportunity to improve on the best known convergence ratio, the algorithm becomes primarily exploratory and favours points $\boldsymbol{x}$ with the largest model uncertainty. For intermediate cases, the strategy described in Eq. (6.2) balances exploitation (maximizing μ) and exploration (maximizing σ). This trade-off between exploitation and exploration avoids getting stuck in local maxima: once a region of successful simulations has been well explored, the algorithm samples from a region of larger uncertainty, even if the predicted convergence ratio is mediocre.

6.2 Global optimisation of convergence using active machine learning

In this section, the point $\boldsymbol{x}$ represents the parameters of a GaN transistor simulation (see Fig. 6.1 for a schematic). To demonstrate our method, we optimise 5 parameters of the simulation. Therefore, each input point $\boldsymbol{x}$ has 5 parameters: 3 physical parameters (the polarisation charge of the transistor channel, the polarisation charge of the top side barrier, and the p type trap concentration of the buffer layer) and 2 numerical parameters (the damping step of ramping the gate-source voltage, and the damping step of ramping the drain-source voltage). We chose here to use our algorithm on GaN transistor simulations, as they are among the most difficult to achieve convergence with.

Our simulations of GaN transistor devices rely on the APSYS software package, with materials parameters taken from [Piprek, 2007]. The band structure was computed using the 6 x 6 k.p method [Chuang, 1991] in a finite volume approximation. The equations related to carrier transport were

self-consistently computed and coupled with Schrödinger's equation, in order to determine the confined states in the quantum well formed by the channel. Schrödinger and Poisson equations were solved iteratively to account for the band structure deformation with carrier redistribution. The carrier transport consists of drift-diffusion of electrons and holes, Fermi statistics, and thermionic emission at hetero-interfaces, as well as band-to-band tunneling. In the simulations, we ramped up the drain-source current of the transistor to from 0V to 10V, and then ramped the gate-source current from 0V to -6V.

We use the machine learning algorithm described in Eqs. (6.1) and (6.2) to optimise convergence within the 5-dimensional space of our transistor simulations. Fig. 6.2 illustrates the strength of our algorithm: the first 100 iterations show the convergence of simulations with random parameters, while the following simulations are automatically set up through our random-forest based active learning strategy. Once this strategy kicks in, it is able to find mostly converged simulations in about 10 iterations. Subsequent iterations lead to little improvement in convergence, and focus instead on decreasing model uncertainty. Learning steps 150 through 700 (Fig. 6.2) allow to construct a very robust model over the global space of transistor simulation parameters. At iteration 700 the algorithm is fully converged, and the coefficient of determination on our testing set is $R^2 > 0.98$ (Fig. 6.4).

6.3 Conclusion

In summary, our active learning strategy rapidly identifies physical and/or numerical parameters leading to the convergence of advanced GaN transistor simulations, while simultaneously building a random forest regression model that predicts the convergence of a wide range of structures. We relied on the objective function in (6.1) for experimental design, which balances exploitation (high predicted efficiency) and exploration (high model uncertainty). At each iteration in our algorithm, this objective function guides the selection of a new transistor simulation, that is used to expand our RF model.

The algorithm we introduce demonstrates the power of machine learning for automated convergence; this method also applies to the convergence of different optoelectronic simulations, such as LEDs.

In this chapter, we demonstrated that global optimisation based on active learning can rapidly and automatically explore Poisson-Schrödinger simulations of gallium nitride devices, and can drastically accelerate their convergence, a process that can take weeks of work to achieve by hand.

Chapter 7

Machine learning for the automatic design and optimisation of LEDs

The use of GaN-based light emitting diodes (LEDs) for very high light output applications is limited by their so-called efficiency droop [Stevenson, 2009]. Beyond high current densities (>10A/cm^2), LED efficiency at room temperature quickly drops as a function of injected current. There is an ongoing debate regarding the causes of this droop, the main proposed mechanisms being: 1) Auger recombination [Gardner et al., 2007, Iveland et al., 2013]; 2) hole injection efficiency [Xie et al., 2008, Pope et al., 2003]; 3) carrier escape from the active region [Schubert et al., 2009]; and 4) carrier delocalisation effects [Hangleiter et al., 2007, Oliver et al., 2010, Badcock et al., 2013]. However, all these models link efficiency droop to high carrier concentrations within the active region of the LED [Laubsch et al., 2009, Lin et al., 2012a]. Therefore, to delay the onset of droop with respect to the injected current, one seeks to optimise LED structures in order to spread out evenly carrier concentrations across the active region. Although the Poisson-Schrödinger simulations used here are not fully accurate, the same optimisation strategy will also apply directly to laboratory fabrication of LED structures. Our approach allows us to rapidly build a model mapping LED structure to simulated efficiency, thereby overcoming time-consuming trial and error based simulations. The efficiency at high current densities of our simulated structures optimised with machine-learning exceed those of reference LED structures by close to 40%.

We use statistical and machine learning (ML) techniques to accelerate the LED design process, by identifying new structures to sample (i.e. build or simulate) in a highly targeted way. We rely on the Efficient Global Opti-

misation (EGO) strategy of Jones *et al.* [Jones et al., 1998], that iteratively selects sample points with the aim of maximising expected efficiency improvement while simultaneously accounting for model uncertainty. We build an ML regression model to predict the efficiency for as-yet unseen structures, from a "database" of LEDs with known structure (i.e. number of layers, their composition, doping, and widths) labelled by resulting electro-luminescence internal quantum efficiency. After a new LED structure is sampled, its characteristics and resulting efficiency are added to the database, and the ML model is updated. Instead of purely selecting new LED structures for which the ML model predicts the greatest efficiency, our method also favours structures for which the ML model is uncertain. This approach strikes a more delicate balance between exploration and exploitation, and helps to find a globally optimum LED. More specifically, this trade-off allows to improve the *global* accuracy of the ML model, and therefore minimises the chance of getting stuck in a region of the LED design space for which the efficiencies are only local maxima.

7.1 Global optimisation using Gaussian Processes

Our approach to optimisation relies on two components. First, a Gaussian process (GP) regression predicts LED efficiency. Second, the EGO heuristic selects the sample point maximising the *expected improvement* in efficiency, according to the GP model uncertainty. A key advantage of GP regression is that it includes *uncertainties* along with its predictions. Our work builds upon a long history of applying GP to computer simulations [O'Hagan and Kingman, 1978, Sacks et al., 1989] and to global optimisation [Journel and Huijbregts, 1978, Gupta and Miescke, 1996, Jones et al., 1998, Osborne et al., 2009, Frazier et al., 2009, Srinivas et al., 2012]. Here we work in the space of LED structures $\boldsymbol{x}$, and the GP models the probability distribution of the simulated electro-luminescence efficiency $y = f(\boldsymbol{x})$. GPs can be interpreted in a Bayesian context, in which the characteristics of the function being learned (such as smoothness and variation length-scale) are modeled as a 'prior' distribution over the general space of functions. Then, given a dataset $\mathcal{D}$ of LED structures and their associated efficiencies, the Bayes rule can be used to obtain a 'posterior' normal distribution for LED efficiency [Rasmussen, 2006]. More specifically, for an as-yet unobserved LED structure $\boldsymbol{x}$, the GP model produces a posterior distribution $P(y|\boldsymbol{x}, \mathcal{D}) \propto \exp[(y-\mu)^2/2\sigma^2]$, where μ corresponds to the expected efficiency and σ to its standard devi-

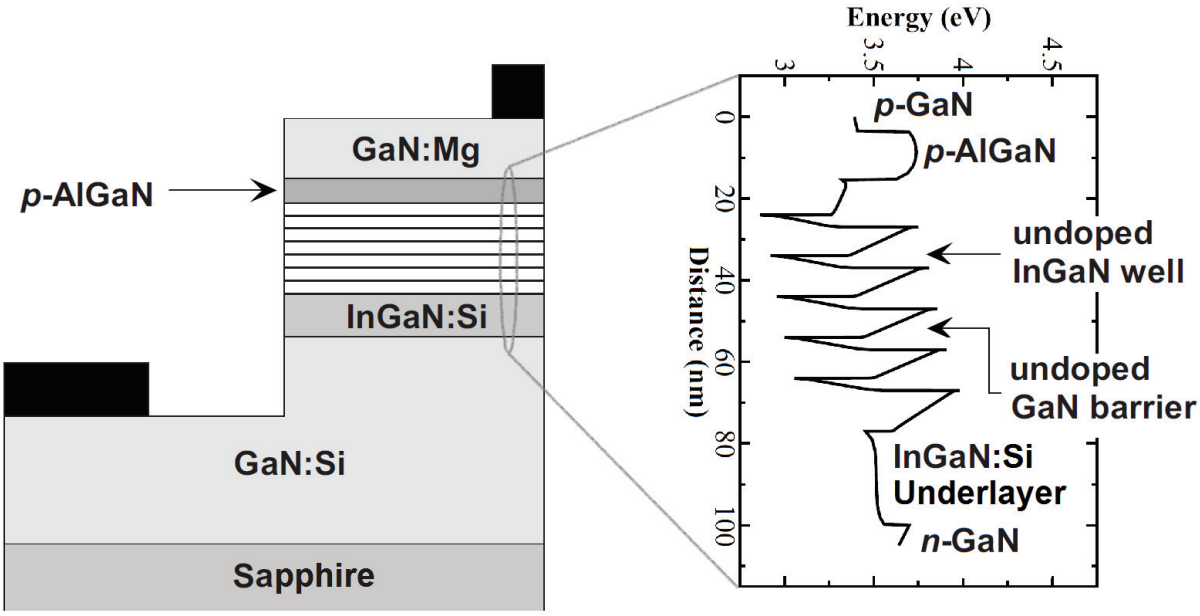
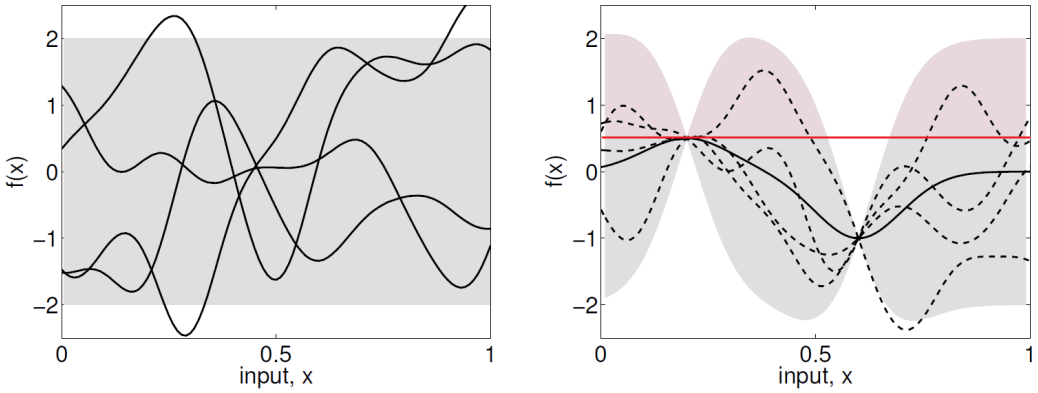


Figure 7.1: Schematic of the LED that the learning algorithm was given the task to optimise. Left: reference LED structure. Right: conduction band of the active region of the reference LED structure, at high current density ($75\text{A}/\text{cm}^2$).

ation. In this chapter, we use the `gaussian_process` module of the scikit-learn Python package [Pedregosa et al., 2011, Lophaven et al., 2002], with a squared-exponential auto-correlation function and hyper-parameters determined by the maximum likelihood principle.

We rely on a dataset $\mathcal{D}_n = \{(\boldsymbol{x}_i, y_i)_{i=1...n}\}$ of LED structures and their measured efficiencies. For a LED structure $\boldsymbol{x}$, the GP model produces a normal posterior distribution $P(y|\boldsymbol{x}, \mathcal{D}_n)$ with mean efficiency μ and standard deviation σ. We aim at identifying new LED structures $(\boldsymbol{x}_{n+1}, \boldsymbol{x}_{n+2}, ...)$ such that we most quickly find LEDs with very high efficiencies. For this purpose, we rely on the EGO strategy: at each step in the design loop, the next LED structure to sample should be selected to *optimise the expected improvement in efficiency*, after accounting for model uncertainty [Jones et al., 1998]. More specifically, with $y_{\max}$ denote the efficiency of the best LED device currently in our dataset $\mathcal{D}_n$, the expected efficiency improvement may be expressed as Eq. 6.1.

Fig. 7.2b provides a visualisation of the EGO strategy: the next simulation $\boldsymbol{x}$ to be run is the one that corresponds to the largest red area, i.e.



(a), prior (b), posterior

Figure 7.2: Adapted from [Rasmussen, 2006]. Gaussian processes-based models have built-in confidence intervals, that we use to select the most promising simulation to run. (a): The prior distribution of possible proxy functions of the simulator. (b): The posterior distribution of proxy functions, filtered to account for the observed results of the simulator. The red area represents the EGO heuristic, that chooses the next simulation based on whether it is highly promising or highly uncertain. Where the standard deviation of the posterior model reaches 0, it indicates a data point.

to the largest possible improvement. Note that in this theoretical example figure, the algorithm already found the best possible structure, but will still explore structures with great uncertainty.

We select the next sample point to optimise the following objective function:

$$\boldsymbol{x}_{n+1} = \arg\max_{\boldsymbol{x}} v(\boldsymbol{x}, \mathcal{D}_n). \tag{7.1}$$

The new LED structure $\boldsymbol{x}_{n+1}$ becomes an input to a Poisson-Schrödinger code (described below) calculating the simulated efficiency y_{n+1}. Our dataset $\mathcal{D}_{n+1} = \mathcal{D}_n \cup \{(\boldsymbol{x}_{n+1}, y_{n+1})\}$ is then extended, and we update the GP posterior $P(y|\boldsymbol{x}, \mathcal{D}_{n+1})$, which is used to select *another* sample point via Eq. (7.1). This iterative process is repeated until an efficient LED structure is found. At the same time, this approach builds a predictive ML model of LED efficiency over a broad range of inputs.

To reach a better understanding of the objective function in Eq. (6.1), we evaluate it within two asymptotic limits. In the limit of vanishing uncertainty, $\sigma \to 0$ (or equivalently $\alpha \to \pm\infty$), we observe $v(\boldsymbol{x}, \mathcal{D}_n) \to (\mu - y_{\max} + |\mu - y_{\max}|)/2$. Therefore, when the model is very certain, the objective function seeks primarily to select points $\boldsymbol{x}$ with higher expected efficiency μ

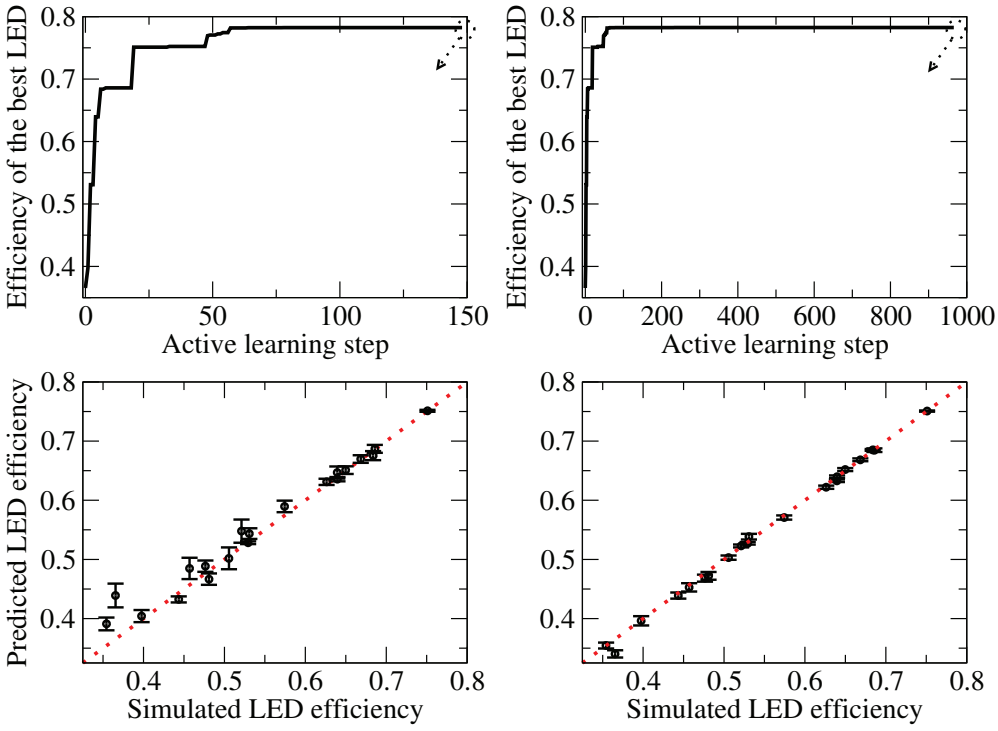


Figure 7.3: Top row: Simulated room temperature EL internal quantum efficiency of the best LED structure known at each learning step. Bottom row: Efficiencies predicted by machine learning versus simulated efficiencies for structures unseen by the algorithm (out of sample) after 150 (left) and 1000 (right) iterations. This testing set consists in 20 simulations with random parameters. The input of the model consists in the 6 structural parameters described in the text, and the output is the predicted simulated efficiency. The input structural parameters are allowed to vary in ranges that span 1 order of magnitude. The database is built by running the Poisson-Schrödinger simulations. First randomly, to build the testing set, and then according to parameters chosen by the GP model once the active learning starts and incrementally builds the training set. The error bars indicate the uncertainty of the machine learning model, as given by the GP model. The active learning strategy first focuses on high efficiencies, and then improves its model for the lower efficiency structures, which is reflected in evolution of the error bars with active learning step.

than the best known $y_{\max}$. In contrast when the model's uncertainties σ are relatively large compared to $\mu - y_{\max}$, i.e. $\alpha \to 0$, we observe $v(\boldsymbol{x}, \mathcal{D}_n) \to \frac{\sigma}{\sqrt{2\pi}}$.

This suggests that when there are no obvious opportunities to improve on the best known LED, the learning strategy becomes primarily exploratory and favours points $\boldsymbol{x}$ with the largest model uncertainty. For intermediate cases ($\alpha \sim 1$), the strategy of Eq. (7.1) balances exploitation (maximizing μ) and exploration (maximizing σ). This allows to avoid getting stuck in local maxima: once a region of very efficient LEDs has been explored well enough, the algorithm samples from a region of larger uncertainty - even if the predicted efficiency is not very high.

7.2 Optimisation of GaN LEDs and the reduction of efficiency droop using active machine learning

In this chapter, a point $\boldsymbol{x}$ represents the structure of the 5-well active region in a GaN-based LED (see Fig. 7.1 for a schematic). Each input point $\boldsymbol{x}$ has 6 parameters, that are the indium composition of each quantum well and the collective indium composition of the quantum barriers. The width of the quantum well varies with the indium composition of both well and barrier, in order to maintain the wavelength approximately constant. We use the APSYS software package (with materials parameters taken from [Piprek, 2007] and current density 75A/cm^2) to determine the simulated efficiency of each structure, . The band structure was calculated with the 6 x 6 k.p method [Chuang, 1991] in a finite volume approximation. The carrier transport equations were self-consistently computed and coupled with Schrödinger's equation to determine the confined states in the QWs. Schrödinger and Poisson equations are solved iteratively to account for the band structure deformation with carrier redistribution. The carrier transport consists of drift-diffusion of electrons and holes, Fermi statistics, and thermionic emission at hetero-interfaces, as well as band-to-band tunneling.

We use the machine learning algorithm described in Eqs. (5.1) and (7.1) to optimise the internal quantum efficiency within the 6-dimensional space of our LED structures (the In content of each the 5 wells, and the average In content of the barriers). As can be seen in Fig. 7.3**a**, the algorithm converges rapidly, and finds a nearly optimal simulated LED efficiency in about 75 iterations. Subsequent iterations make little improvement regarding LED efficiency (Fig. 7.3**b**), and concentrate instead on decreasing the model's uncertainty. Between learning steps 150 through 1000 (Fig. 7.3**c** and **d**), the algorithm constructs a very robust model over the global space of LED structures. At iteration 1000 the algorithm has converged, and the coefficient

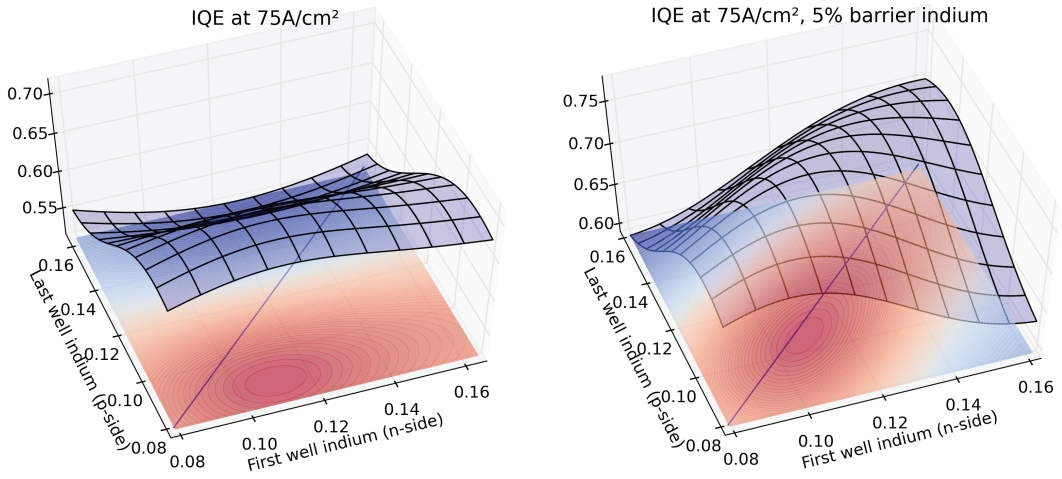


Figure 7.4: Gaussian process model of expected LED efficiency as a function of the indium content of each well in the active region, linearly interpolated from the n-side (first well) to the p-side (last well). The model is built upon 1000 APSYS simulations spawned by the active learning algorithm, and predicts LED efficiency with near perfect accuracy. Left: No indium in the barriers. Right: 5% indium in the barriers. In both cases the optimum is reached with the indium content of the wells decreasing from the n-side to the p-side. In our iterative active learning strategy, the width of the simulated quantum well is adjusted to keep the emission wavelength constant.

of determination is $R^2 > 0.99$ (as determined by cross-validation).

The very high accuracy of the model provides some insight into the physics of the Poisson-Schrodinger simulations. The drift-diffusion model predicts that majority of the light emission of a standard LED structure comes from the two top wells, in agreement with electro-luminescence experiments [David et al., 2008]. It also informs us that varying indium contents of individual wells across the active region increases the carrier and light emission spreading, in agreement with recent electro-luminescence experiments [Kang et al., 2015]. As can be seen in Fig. 7.4, our active learning algorithm finds several optima. All these optima have in common a decreasing indium content in the quantum wells from the n-side to the p-side, and the use of InGaN barriers rather than GaN barriers. The decreasing indium content reduces the confinement in the p-side wells [Sze and Ng, 2006], that otherwise concentrate most carriers. The decreasing indium content and the use of InGaN

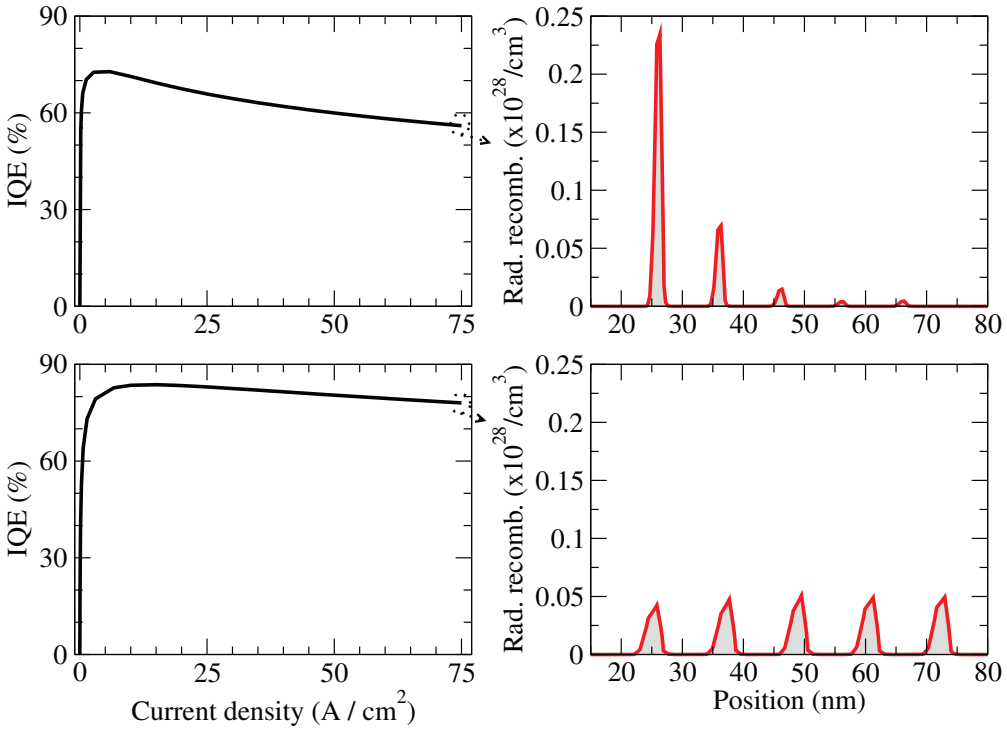


Figure 7.5: Comparison between the simulation of an initial standard LED structure (top row) and the simulation of an LED structure optimised by machine learning (bottom row). Left: EL Internal quantum efficiency (IQE) as a function of current density. Right: Radiative recombination rate through the active region, at a current density of 75A/cm^2. The p-side is towards the left. The optimised structure has wider wells, that are getting shallower (less In) towards the p-side.

quantum barriers increases the thermionic emission and tunnelling through the hetero-interfaces [Grinberg et al., 1984], allowing the carriers to spread more easily across the active region.

The decreasing indium content with higher well number is associated with increasing well widths for a constant peak emission wavelength. At high current, Auger recombination increases more rapidly with carrier concentration than radiative recombination, and wider wells that compensate for the low indium content become beneficial [Li et al., 2012], as the carrier spreading within each well is increased. Fig. 7.5 draws a comparison between the simulation of a standard reference LED structure that has GaN barri-

ers and identical wells with the simulation of a machine learning optimised LED structure. The optimised structure achieves a better spreading of the radiative recombination events: within the wells due to wider wells, which should be beneficial at high currents [Ren, 2015], and between the wells due to a high barrier indium content and a decreasing indium content towards the p-side of the active region.

7.3 Conclusion

To conclude, our active learning strategy quickly finds LED structures with nearly optimal quantum efficiency while building at the same time a GP regression model that is predictive for a wide range of LEDs. We used the objective function in (5.1) for active learning, which balances the trade-off between exploitation (high predicted efficiency) and exploration (high model uncertainty). At each iteration in our algorithm, the objective function guides the selection of a new LED structure which we simulate, and then use to improve our GP model.

Interestingly, this automated approach finds LEDs that a human expert would want to design: a structure that spreads more evenly the carrier recombination events through the active region of the LED, maximising the radiative recombination events. Leaving the algorithm to optimise the indium content of the active region, we find dramatically increased simulated efficiencies than in standard LEDs. This structure employs a high barrier indium content and a well indium content that decreases towards the p-side of the active region to prevent the accumulation of carriers on the p-side and improve the spreading of the carriers and the radiative recombination events. Additionally, it employs wider wells to compensate for wavelength changes with indium content, and to achieve a carrier spreading within the quantum wells that is desirable at high currents.

Our calculations of gallium nitride devices with Poisson-Schrödinger solvers provide qualitative information rather than quantitative predictions. Nevertheless, the algorithm we present demonstrates the power of machine learning for device design. Our method also applies to the optimisation of different LED structures than those presented here. Coupled with the automatic parametrisation of simulations on experimental results (Chap. 5), it still provides predictions that are not strictly theoretical, but based on experiments. When used in conjunction with actual materials fabrication, our method could extend to the design of experimental devices.

Materials informatics is an emerging field [Rajan, 2005, Jain et al., 2013, Balachandran et al., 2015] with great promise for new materials design [Bhadeshia,

1999, Hautier et al., 2010, Pilania et al., 2013, Castelli and Jacobsen, 2014, Ghiringhelli et al., 2015]. This approach has not yet been used by the LED community, despite great potential for improving physical understanding and for accelerating the design of devices. In this work, we show that active learning based global optimisation can rapidly and automatically explore Poisson-Schrödinger simulations of gallium nitride devices, and can accelerate the discovery of efficient LEDs.

Chapter 8

Machine learning predicts material failure in laboratory earthquake experiment

Forecasting fault failure is a fundamental but elusive goal in earthquake science. We apply machine learning to data sets from shear laboratory experiments, with the goal of identifying hidden signals that precede earthquakes. Here we show that by listening to the acoustic signal emitted by a laboratory fault, machine learning can predict the time remaining before it fails with great accuracy. These predictions are based solely on the instantaneous physical characteristics of the acoustical signal, and do not make use of its history [Rouet-Leduc et al., 2017]. Surprisingly, machine learning identifies a signal emitted from the fault zone previously thought to be low-amplitude noise that enables failure forecasting throughout the laboratory quake cycle. We hypothesize that applying this approach to continuous seismic data may lead to significant advances in identifying currently unknown signals, in providing new insights into fault physics, and in placing bounds on fault failure times.

A classical approach to determining that an earthquake may be looming is based on the inter-event time (recurrence interval) for characteristic earthquakes-earthquakes that repeat periodically [Schwartz and Coppersmith, 1984]. For instance, analysis of turbidite stratigraphy deposited during successive earthquakes dating back 10,000 years suggests the Cascadia subduction zone is ripe for a megaquake [Goldfinger et al., 2016] (Fig. 8.1). The idea behind characteristic, repeating earthquakes was the basis of the well-known Parkfield prediction based strictly on seismic data. Similar earthquakes between 1857 and 1966 suggested a pattern of quakes every 21.9±3.1 years. Based on the recurrence interval an earthquake was expected between

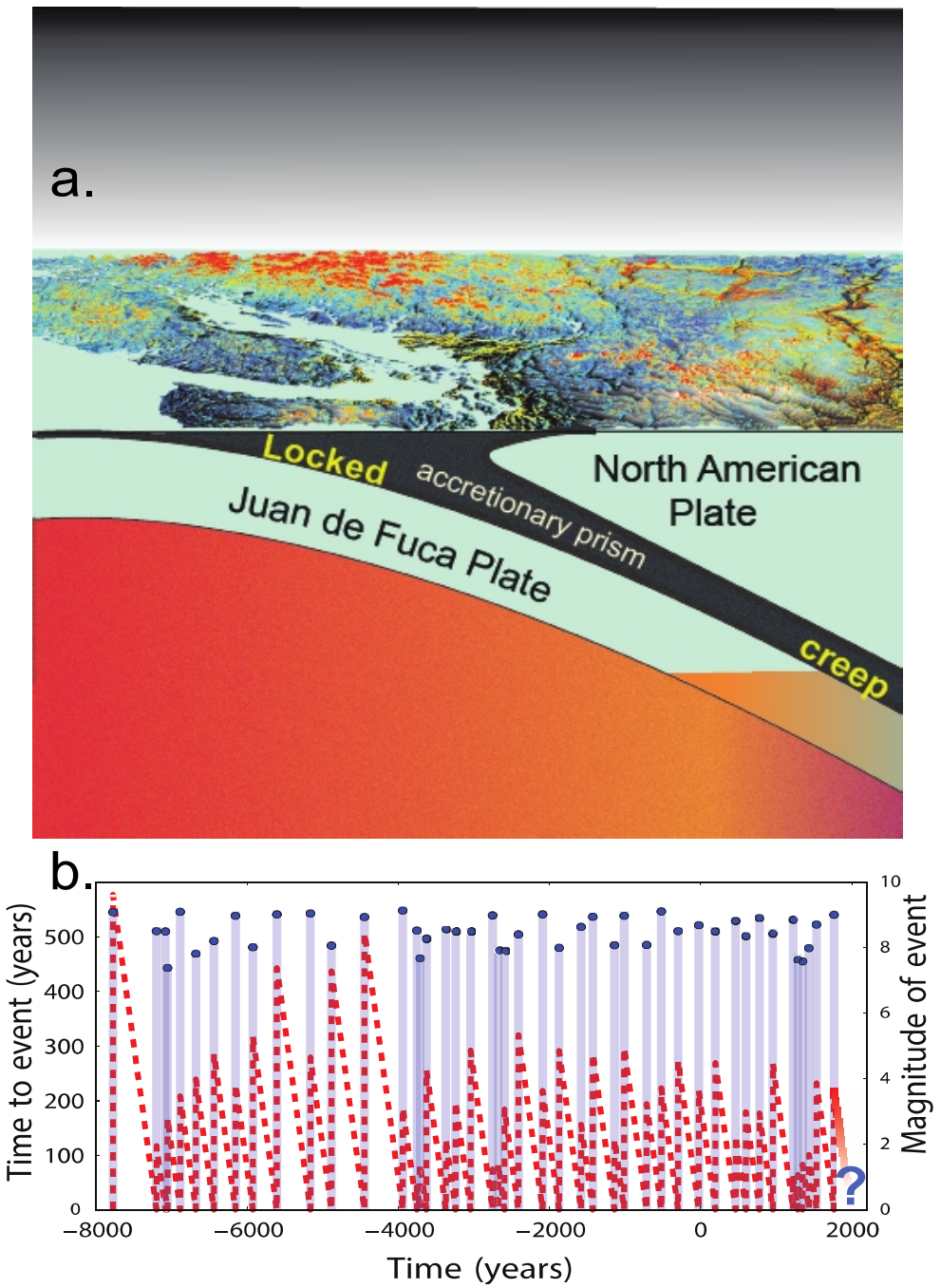


Figure 8.1: Subduction in Cascadia and the large earthquakes that have occurred in the last 10,000 years. **a**: Illustration of subduction of the Juan de Fuca Plate subducting beneath the North American plate in the vicinity of Seattle. **b**: Earthquakes >M8.0 estimated from associated oceanic landslides, called turbidites [Goldfinger et al., 2016]. The blue dots with gray vertical lines show earthquake magnitude estimates with calendar time, and the red stippled line shows the time remaining before the next event. Cascadia is locked and stressed, it is currently due for a megaquake and accompanying tsunami..

1988-1993 [Bakun and Lindh, 1985], but ultimately took place in 2004. With this approach, as earthquake recurrence is not constant for a given fault, event occurrence can only be inferred within large error bounds.

Over the last 15 years, there has been renewed hope that progress can be made regarding forecasting owing to tremendous advances in instrumentation quality and density. These advances have led to exciting discoveries of previously unidentified slip processes that include slow slip [Melbourne and Webb, 2003], Low Frequency Earthquakes and Earth tremor [Shelly et al., 2007, Obara, 2002] that occur deep in faults. These discoveries inform a new understanding of fault slip and may well lead to advances in forecasting impending fault failure if the coupling of deep faults to the seismogenic zone can be unraveled.

The advances in instrumentation sensitivity and density also provide new means to record small events that may be precursors. Acoustic/seismic precursors to failure appear to be a nearly universal phenomenon in materials. For instance, it is well established that failure in granular materials [Michlmayr et al., 2013] and in avalanche [Pradhan et al., 2006] is frequently accompanied by impulsive acoustic/seismic precursors, many of them very small. Precursors are also routinely observed in brittle failure of a spectrum of industrial [Huang et al., 1998] and Earth materials [Schubnel et al., 2013, Jaeger et al., 2009]. Precursors are observed in laboratory faults [Johnson et al., 2013, Goebel et al., 2013] and are widely but not systematically observed preceding earthquakes[Bouchon et al., 2013, Bouchon et al., 2016, McGuire et al., 2015, Mignan, 2014, Wyss and Booth, 1997, Geller et al., 1997]. The International Commission on Earthquake Forecasting for Civil Protection concluded in 2011 there was "considerable room for methodological improvements in this type of [precursor-based failure forecasting] research" [Jordan et al., 2011]. The commission also concluded that published results may be biased toward positive observations. We hypothesize that precursors are a manifestation of critical stress conditions preceding shear failure. We posit that seismic precursor magnitudes can be very small and thus frequently go unrecorded or unidentified. As instrumentation improves, precursors may ultimately be found to exist for most or all earthquakes [Delorey et al., 2017]. Furthermore, it is plausible that other signals exist that presage failure.

8.1 Stick-slip failure in sheared granular material

To address these questions we analyze data from a laboratory fault system, a two-fault configuration that contains fault gouge material submitted to double direct shear (Fig. 8.5). An accelerometer records the acoustic emission (AE) emanating from the shearing layers. The shear stress imposed by the driving block is also monitored (Figs. 8.1, 8.2a, 8.5), as well as other physical parameters such as the shearing rate, gouge layer thickness, friction and the applied load [Scholz et al., 1973]. Following a frictional failure (labquake) the shearing block displaces while the gouge material simultaneously dilates and strengthens, as manifested by increasing shear stress (Fig. 8.2a) and friction. As the material approaches failure, it exhibits characteristics of a critical stress regime, including many small shear failures that emit impulsive AEs [Johnson et al., 2013, Ferdowsi et al., 2015]. This unstable state concludes with a labquake, in which the shearing block rapidly displaces, the friction and shear stress decrease precipitously due to the gouge failure (Fig. 8.5, inset), and the gouge layers simultaneously compact. Under a broad range of load and shear velocity conditions, the apparatus slide-slips quasi-periodically for hundreds of stress cycles during a single experiment (Fig. 8.5) [Marone, 1998, Johnson et al., 2016]. The rate of impulsive precursors accelerates as failure approaches [Johnson et al., 2013], suggesting that upcoming labquake timing could be predicted. Notably, the Gutenberg-Richter scaling [Gutenberg and Richter, 1949] relation calculated from the precursors [Johnson et al., 2013] is within the bounds observed in Earth, implying that some of the important fault frictional physics may scale. In this work, we ask: can the failure time of an upcoming labquake be predicted using characteristics of the continuously recorded acoustic signal?

8.2 Building features from the experimental signal

To solve this problem we apply a machine learning technique, the random forest (RF) [Breiman, 2001], to create a model that forecasts labquakes exclusively from the continuous acoustic time series data recorded from the fault zone (see Fig. 8.2 and the implementation section for full details). Figure 8.2a, top, shows the laboratory shear stress exhibiting multiple failure events during an experiment. Our goal is to predict the time remaining before the next failure (Fig. 8.2a, bottom) using only local, moving time windows of

the AE data (Fig. 8.2b, top). From each window, we compute a set of approximately 100 potentially relevant statistical features (e.g. mean, variance, kurtosis, autocorrelation, etc.) that the RF uses to predict the time remaining before the next failure. Fig. 8.2b shows four of these features on the same time scale as in Fig. 8.2a, through multiple failure cycles. Some features are sensitive to changes in signal characteristics early in time during the stress cycle, just following a labquake. All features shown are strongly sensitive to signal characteristics just preceding failure, as the system approaches shear-stress criticality.

8.3 Predicting the physical behaviour of the system from a tiny sample of real-time signal

The RF model is an average over a set of decision trees (Fig. 8.2c). Each decision tree predicts the time remaining before the next failure using a sequence of decisions based on the statistical features derived from the time windows (see SI Materials and Methods for details). Applying these and other features, we train the RF to predict upcoming slip events simultaneously observed in the shear stress signal (Fig. 8.2a). Fig. 8.2d and 8.3 show failure predictions on testing data, a sequence of slip events not used while generating the model. The red dashed line shows the time remaining before the next failure (derived from the shear stress data), and the blue line shows the corresponding prediction of the RF model (derived exclusively from the 'instantaneous' acoustic data). The blue shaded region indicates the 5th and 95th percentiles of the forecast, that is, 90 percent of the trees that comprise the forest made a forecast within these bounds. We emphasize that there is no past or future information considered in the prediction made from successive time windows. By listening to the acoustic signal currently emitted by the system, we predict the time remaining before it fails. These predictions are thus based on the instantaneous physical characteristics of the system and do not make use of its history.

Figure 8.3 shows in more detail the RF predictions. The time to failure predictions are highly accurate, with an R2 value of 0.89. In fact, the RF model accurately predicts failure throughout the entire labquake cycle, demonstrating that the system continuously progresses towards failure through the entire stress cycle. We find that statistics quantifying the signal amplitude distribution (e.g. its variance and higher order moments) are highly effective at forecasting failure. The variance, which characterizes

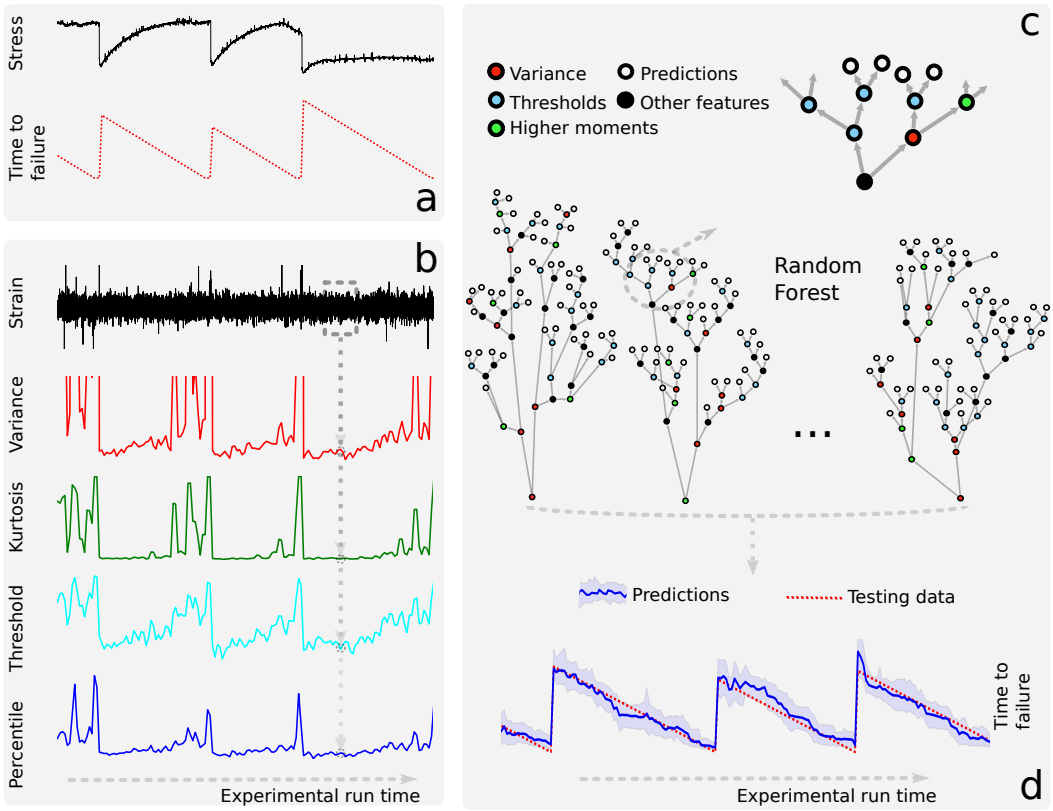


Figure 8.2: Random Forest (RF) approach for predicting time remaining before failure. **a**: Shear stress (black curve) exhibits sharp drops, indicating failure events (labquakes). We wish to predict the time remaining before the next failure derived from the shear stress drops (red curve), using only the acoustic emission (dynamic strain) data (b). **b**: The dashed rectangle represents a moving time window; each window generates a single point on each feature curve below (e.g., variance, kurtosis, etc.). **c**: The RF model predicts the time remaining before the next failure by averaging the predictions of 1000 decision trees for each time window. Each tree makes its prediction (white leaf node), following a series of decisions (colored nodes) based on features of the acoustic signal during the current window (see Supplementary Materials). **d**: The RF prediction (blue line) on data it has never seen (testing data) with 90% confidence intervals (blue shaded region). The predictions agree remarkably well with the actual remaining times before failure (red curve). We emphasize that the testing data is entirely independent of the training data, and was not used to construct the model.

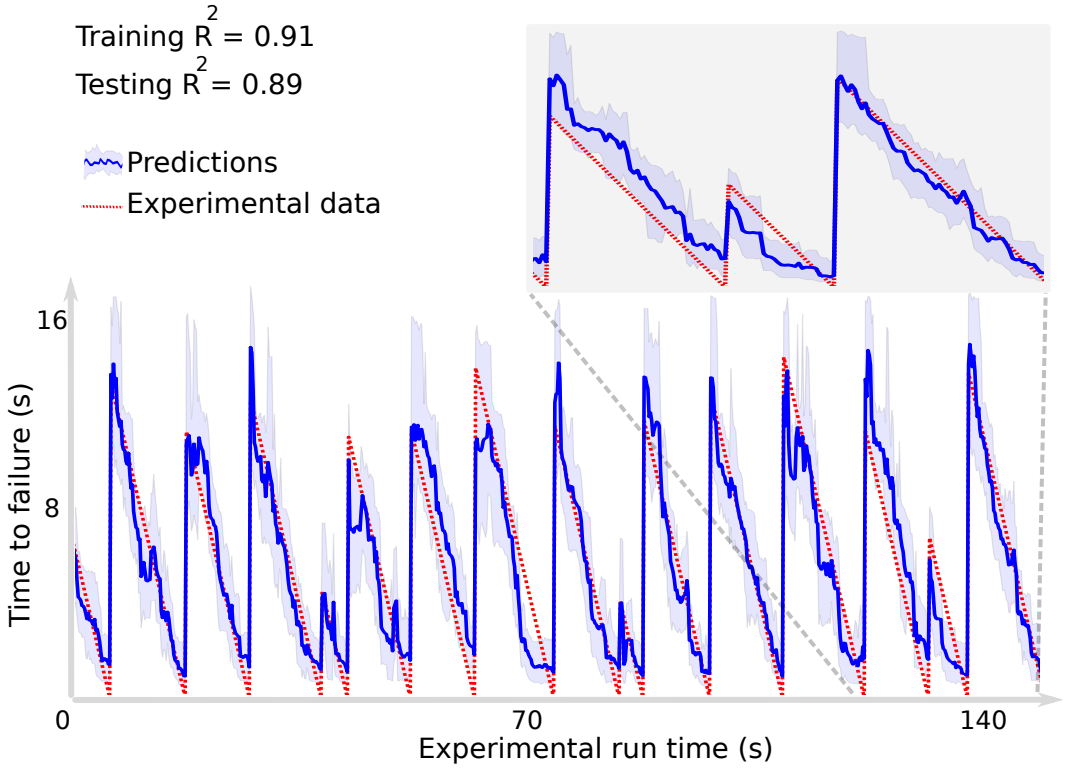


Figure 8.3: Time remaining before the next failure predicted by the Random Forest. As in Fig. 8.2 a and d, the red lines show the actual time before failure (Y-axis) versus experimental run time (X-axis). The blue solid line shows the prediction from the forest, obtained from successive time windows. The shaded region shows the 5 and 95 percentile, 90 percent of the trees that compose the forest provide a forecast within these bounds. The inset emphasizes predictions on aperiodic slip behavior. The RF does a remarkable job of forecasting slip times even with aperiodic data. The RF was trained on 150 seconds of data (10 slip events), and tested on the following 150 seconds, shown here. We stress that the predictions from each time are entirely independent of past and future history, each blue point is a 'now' prediction.

overall signal amplitude fluctuation, is the strongest single feature early in time (Fig. 8.2b). Remarkably, immediately after a failure occurs, the variance accurately predicts the time remaining before the next failure event. As the system nears failure, outlier statistics become the best predictors of failure and best identify impulsive precursors previously associated with critical

stress [Johnson et al., 2013].

8.4 The machine learning model discovers a new signal in the experimental data

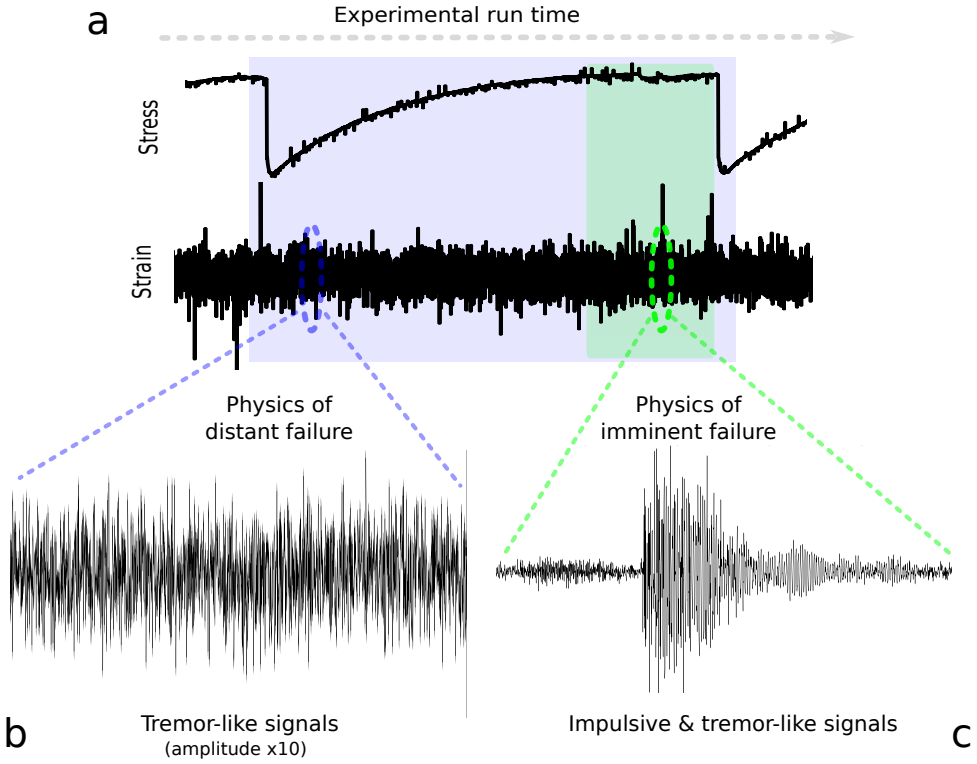


Figure 8.4: The physics of failure. The RF identifies two classes of signals and uses them to predict failure. **a**: Shear stress and dynamic strain encompassing two failure events. **b**: Zoom of dynamic strain when failure is in the distant future. This newly identified signal, termed 'laboratory tremor' offers precise predictive capability of the next failure time. **c**: Zoom of a classic, impulsive acoustic emission observed in the critically stressed region just preceding failure (note vertical scale is different for two signals). Such signals are routinely identified preceding failure in the shear apparatus, in brittle failure in most materials and in some earthquakes.

Our machine learning analysis provides new insight into the slip physics. Specifically, the AE signal occurring long before failure (from the start of

the stress cycle) was previously assumed to be noise and thus overlooked [Johnson et al., 2013]. Fig. (8.4b) shows a raw time series far from failure. The signal exhibits small modulations challenging to identify by eye that persist throughout the stress cycle. The signal resembles volcanic [Chouet, 1996] (harmonic) and non-volcanic [Shelly et al., 2007, Obara, 2002] tremor that exhibit similar ringing characteristics over long periods of time. Of significance is that non-volcanic tremor in fault zones is associated with slow slip [Rogers and Dragert, 2003, Rubinstein et al., 2009]. In the laboratory experiments, the central block (Fig. 8.5) slowly slips throughout the stress cycle, briefly accelerating at the time of failure. Thus, we view the laboratory slow slip as analogous to slow slip on Earth faults. We posit that their laboratory tremor-like characteristics are due to 'creaking' from continuous grain motions of the fault gouge due to slow slipping of the block. Ongoing Discrete Element Modeling of this system [Ferdowsi et al., 2014] are being used to identify the precise origin of these signals.

As failure is approached, the variance of the laboratory tremor increases (Fig. 8.2b). Simultaneously, the rare event statistics (e.g. extreme percentiles and counts over thresholds) become strong indicators of failure, as mentioned above (Fig. 8.2b). These indicators are responding to impulsive AE (Fig. 8.4c) typically observed in material failure [Huang et al., 1998, Jaeger et al., 2009], including those under shear conditions in the laboratory [Johnson et al., 2013] and in Earth [Bouchon et al., 2013, Bouchon et al., 2016]. Such signals occur dominantly in the critical stress regime. In contrast to the laboratory tremor, these signals are due to small observable shear failures within the gouge immediately preceding the labquake [Johnson et al., 2013]. These results suggest that the commonly applied approach to failure prediction based exclusively on the analysis of impulsive precursors is incomplete. The predictions of our model generalize across experimental conditions. To demonstrate this, we trained the system at one applied load level, and then tested it on data from different load levels, exhibiting different inter-event times between failures. We observe that the model predictions retain their accuracy across load levels (e.g., Fig. 8.9). Further, when the stress-cycle periodicity is disrupted by a shorter recurrence time as shown in the inset of Fig. 8.3, the RF still does an excellent job in predicting failure time, suggesting that the approach can be generalized to aperiodic fault cycles.

8.5 Implementation of the failure forecasting algorithm

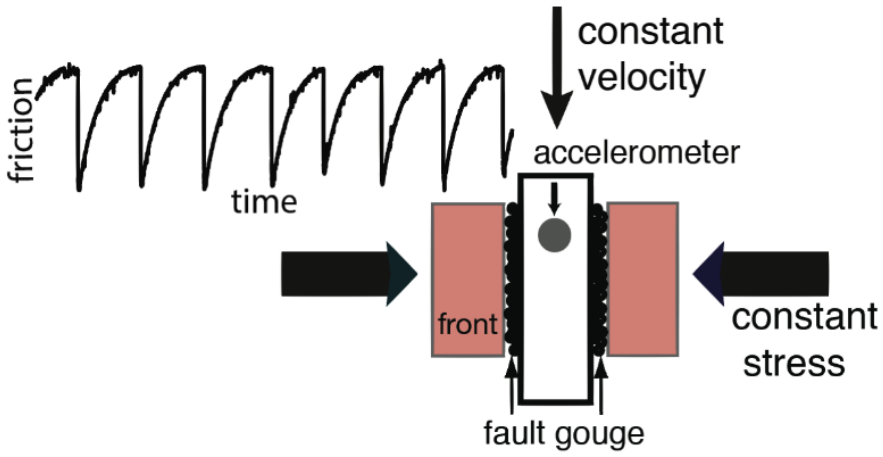


Figure 8.5: Bi-axial shearing device. A slider block with fault gouge layers on either side is loaded by a constant stress. The slider is driven downward at constant displacement rate inducing slide-slip behavior. An accelerometer continuously records the acoustic emission at a sampling rate exceeding 330kHz. Curve (top): recorded friction at gouge interfaces, showing slide-slip events-laboratory earthquakes.

The experimental setup to generate laboratory quakes (Fig. 8.5) has been discussed extensively [Marone, 1998, Niemeijer et al., 2010, Scuderi et al., 2014]. A three-block assembly with two gouge layers is placed in a bi-axial stress configuration. Two 5mm-thick fault gouge layers are placed between the three blocks, which are held in place by a fixed normal load. The gouge material is comprised of Class IV beads with diameter $105 - 149\mu$m. The central block is sheared at constant displacement rate. At some time while the gouge material is in a critical shear stress regime, the shear stress abruptly drops, indicating gouge failure. These large drops in shear stress are laboratory quakes (Fig. 8.5). The two data streams recorded for our purposes here are the shear stress and the acoustic signal. The acoustic particle acceleration $\ddot{u}$ is measured on the central block and can be readily converted to dynamic strain ϵ:

$$\epsilon = \frac{\dot{u}}{c},$$

where $\dot{u} = \ddot{u}/\omega$, $\omega = 2\pi$, with $c \approx 700$m/s the average measured wave speed in the granular material, and $f \approx 40.3$kHz. The sampling rate is 330kHz. As applied load progressively increases, the inter-event time (recurrence) of laboratory earthquakes progressively decreases. At smaller applied loads the slips become aperiodic [Johnson et al., 2013]. In all cases, the rate of impulsive precursors accelerates as failure is approached [Johnson et al., 2013].

8.5.1 Random Forest overview

Machine learning offers a variety of algorithms suitable for modeling the relationship between an input data (here characterized by features derived from a time window of the acoustic emissions signal) and corresponding output label (here the time remaining before the next failure event, derived from the shear stress signal). We give a brief overview of the RF algorithm [Breiman, 2001] used in this study. Details are presented in the Random Forest details and Statistical features sections below.

A trained RF model predicts the time remaining before the next failure, from the features of the input time window. The RF is an ensemble method that makes its prediction by averaging over a set of simple decision trees. The trees are stochastically generated. Although the prediction of an individual decision tree may be quite inaccurate, the errors tend to cancel between trees, and the averaged RF prediction can be remarkably accurate.

A single decision tree (pictured in Fig. 8.2c) operates as follows. To predict the time to failure for a data point extracted from the time series window, we begin at the root node and work toward the leaves. Each internal node encountered represents a binary decision on a single feature of the data point. For example, the decision at the root node could be "Is the variance during this time window greater than 0.5"? If the answer is "yes", one continues up the left branch. Otherwise, one goes up the right branch. This sequence of decisions continues until a leaf node at the top of the tree has been reached (denoted by open circles in Fig. 8.2). Each leaf node contains a possible prediction of the tree, i.e. a predicted time remaining before failure (in seconds) for this particular data point. Note that each leaf node corresponds to a specific sequence of yes/no decisions on the features derived from the data point. Thus, a tree represents a map from inputs (acoustic emission features) to output labels (time remaining before failure prediction).

We build our dataset by computing statistical features from local time windows of the acoustic emission signal. Each data point (i.e. local set of statistical features) is associated with a time to failure, which we calculate from the shear stress signal. We then take our entire collection of data points

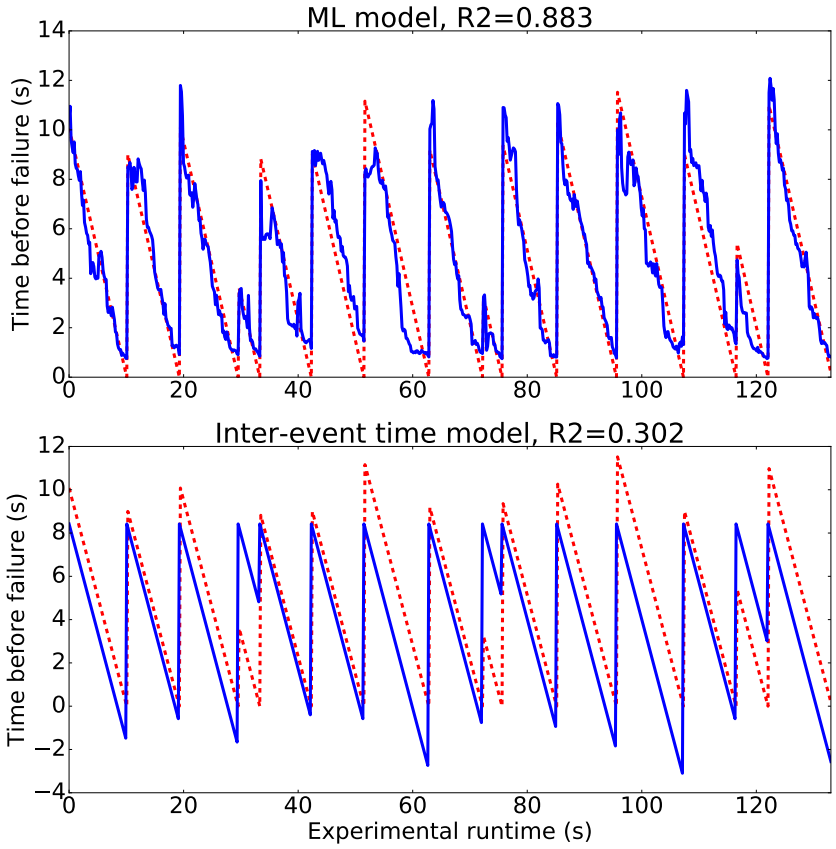


Figure 8.6: Top: the machine learning model. Bottom: a model that uses the average inter-event time of the training data to make predictions. Its performance is terrible: at an $R^2 = 0.3$ it is not much better than a straight line, as it completely misses the outliers. The dashed red curves are the real times remaining before failure, and the blue solid curves are the predictions from the models.

and split it into two disjoint parts: training data and testing data. The training data is used to generate the RF model. The testing data is used to evaluate the performance of the RF model; this constitutes a fair measure of the RF performance because the testing data is independent from the training process (i.e. out of sample performance). It is very important to ensure that testing data does not leak into the training process.

To build the RF, we stochastically generate each decision tree from the training data as follows. The tree is given by a bootstrap resampling of the training data, which induces variation between the trees and mitigates the effect of outliers on the forest. Again, we begin at the root node of the tree and work toward the leaves. To generate each node, we formulate a "yes"/"no" decision (corresponding to left/right branches) operating on the available data. At each node, we select a random subset of 40% of the available features. From the selected features, we construct the decision (i.e. select a feature and threshold value) that best predicts the time to failure. This corresponds to selecting a decision that partitions the current available data into two groups that are maximally dissimilar to each other with respect to time remaining before failure (quantitatively, using a mean squared error splitting criterion). The data is partitioned according to the decision, generating two branches. On each branch, the process repeats recursively, generating the entire tree structure. Decisions are generated until the number of data points at a node has reached a minimum size, at which point the tree constructs a leaf node. The tree assigns a prediction to each leaf node by taking the average time remaining before failure for the data that falls into that leaf.

Note the predictions never reach zero due to the discretization in time of the problem imposed by the moving window approach. In particular, we do not consider the time windows during which a failure occurred (neither for training nor for testing): at the moment failure takes place, all the statistical features are several orders of magnitude higher than the rest of the time. This problem vanishes with smaller windows, at the cost of increased computation.

8.5.2 Random Forest details

For this work, we used the scikit-learn implementation of the random forests [Pedregosa et al., 2011], which implements the algorithm of Breiman25. We compute regularization hyper-parameters by 3-fold cross-validation. The minimum number of samples to generate a split was 30. The minimum number of samples on a leaf was 30. The maximum number of features to consider when making a split was 40 (out of $\approx$ 100 features). The forest size was 1000 trees. The performance of the random forest is not strongly sensitive to this choice of hyper-parameters: changing any hyper-parameter by a factor of 2 typically affects the R2 performance by only a few percent.

To create a model that uncovers the physics of shear failure, we make predictions using moving time windows applied to the data. Each window is 1.8s, which is small compared to the time between fault gouge failures (15s to 60s). The offset between windows is 0.18s, meaning that consecu-

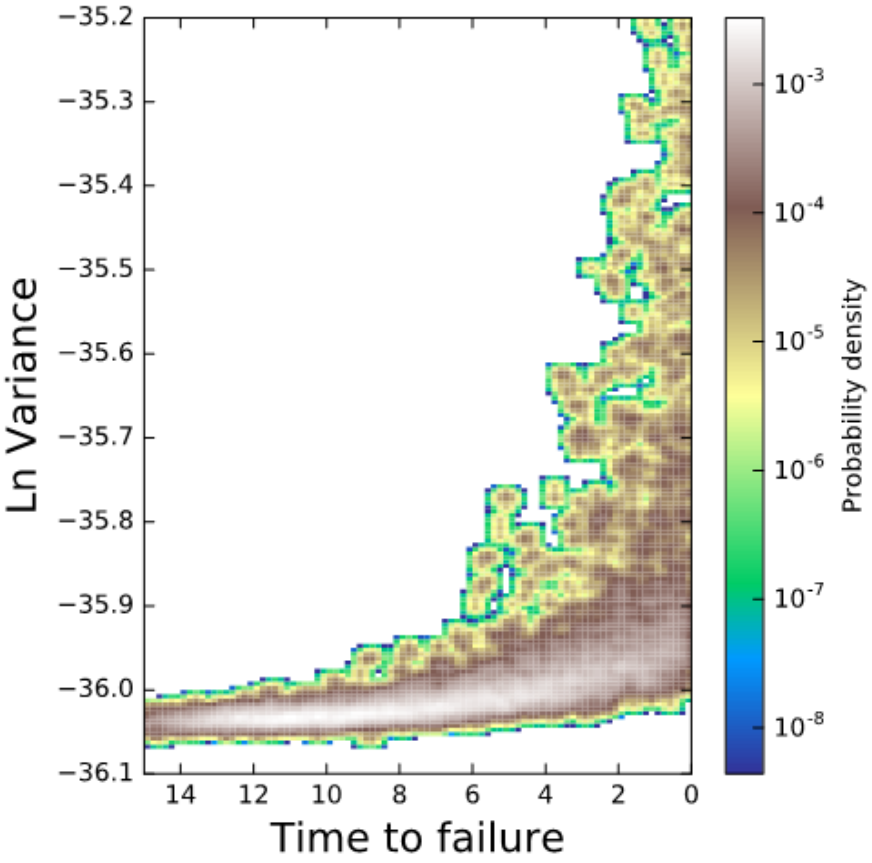


Figure 8.7: The distribution of data points across the variance and the time to failure, shown as a two dimensional histogram. As failure approaches, the amplitude of the variance grows and fluctuates, providing information about the time to failure.

tive time windows overlap by 90%. We characterize the acoustical signal in each window by a set of ≈ 100 of statistical features (detailed in section "Statistical features"). The features from the ith and (i+10)th (adjacent, non-overlapping) time windows are concatenated to form a vector of features x_i. We label the data point x_i according to the time remaining until the next gouge failure, y_i, determined from the stress signal. The machine learning dataset is then $D_n = \{(x_i, y_i)_(i = 1n)\}$. Figure 8.8 shows the random forest learned from two features, the normalized second (variance) and fourth central moments (kurtosis).

We use a 50/50 split of the full time series data into two contiguous pieces

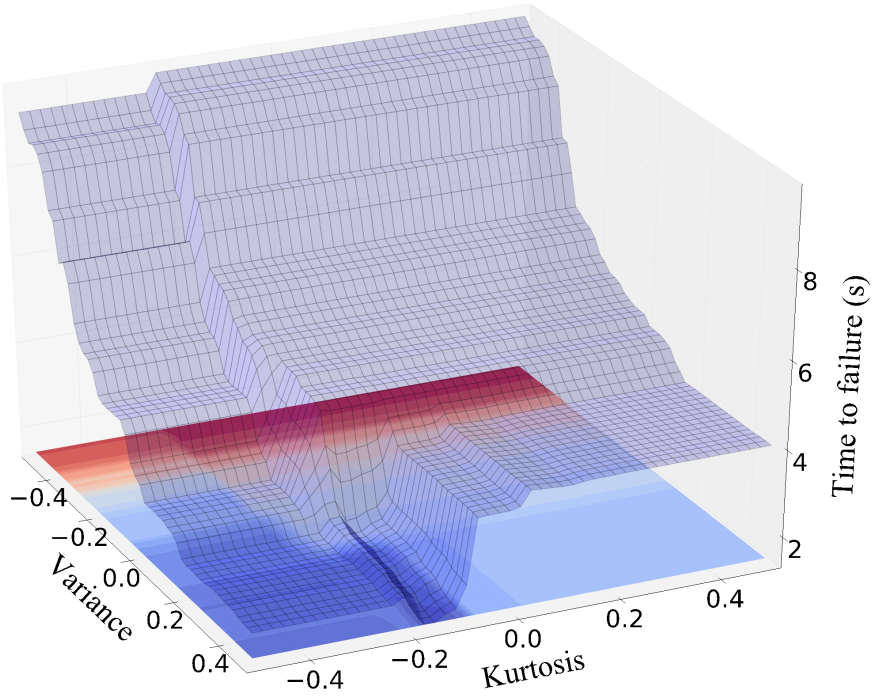


Figure 8.8: The time remaining before the next failure predicted by a RF model constructed from the second and fourth normalized central moments (variance and kurtosis) of the time signal.

for use as training and testing data, respectively. Contiguity of these pieces is important to minimize contamination of the training data with information about the test data, which can arise due to temporal correlations in the full time series. Furthermore, the accuracy of the RF on the training and testing datasets is nearly identical, implying that the model does not overfit the training data, a common concern in machine learning.

8.5.3 Statistical features

We compute many statistical features within each time window for use by the random forest. We computed these features on both the AE signal $f(t) = \epsilon(t)$ and its first finite difference, $f(t) = (\epsilon_(i+1) - \epsilon_i)/(t_(i+1) - t_i) \approx d\epsilon/dt$. Forests that analyze only the derivative of the strain have a slight performance advantage, and so results reported here use only features from

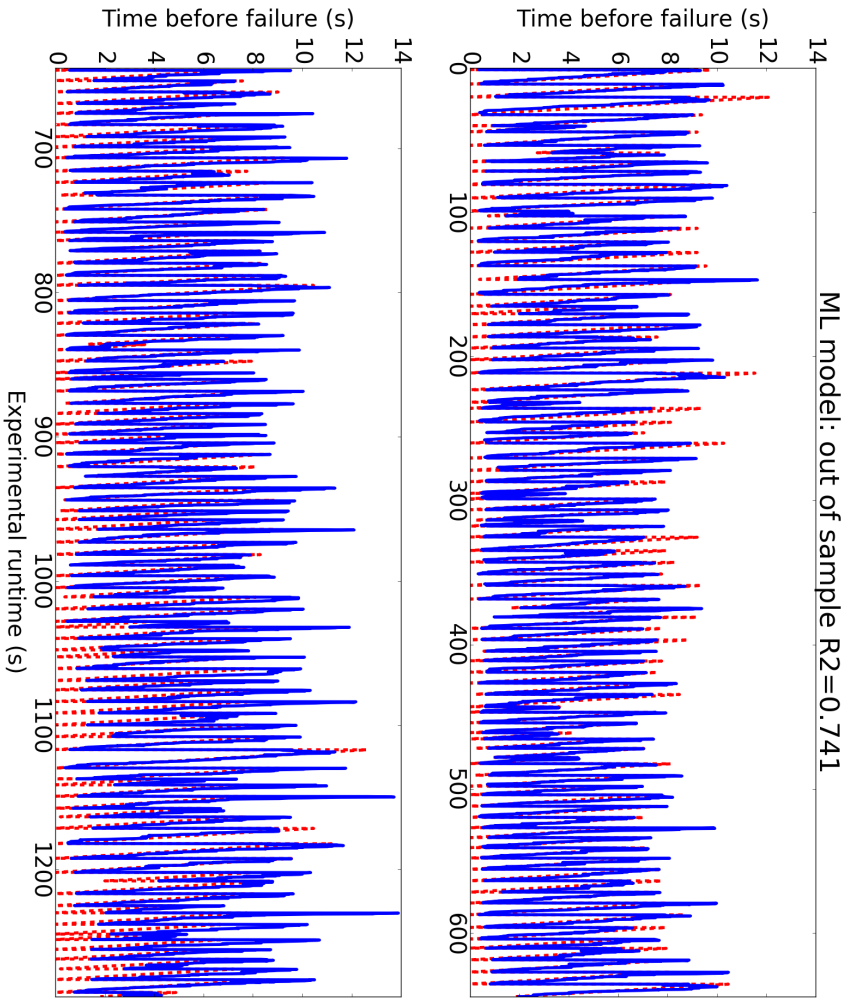


Figure 8.9: Results of training and testing at two different load levels. The dashed red curve is the real time remaining before failure, and the blue solid curve is the prediction from the ML model. Specifically, training was conducted at 5 MPa and testing took place at 8MPa applied load. The in sample R^2 was 0.82 and the out of sample R^2 was 0.741. 'In sample' refers to the training set, and 'out of sample' refers to the testing set. The accuracy is reduced but predictions nonetheless remain highly relevant to new experimental conditions, with clear outliers still being predicted correctly. Where there is a discontinuity of the dashed red curve, or where it does not reach zero, a problem was detected with the acoustic signal, and it was not considered for the analysis (this cleaning takes place before training or testing, when the statistical features are computed).

the derivative signal. The features are as follows:

- Centered moments of orders 1 to 4, with the nth centered moment given by $_n = 1/T \int (t-)^n f(t)dt$, with $= 1/T \int f(t)dt$ the average of $f(t)$ over the time window of length T.
- Signal maximum, signal minimum, and several extreme percentiles. We use the $1st$ to $9th$ and $91th$ to $99th$ percentiles, by increments of 1%.
- Fraction of time $1/T \int \Omega(f(t) - f_0)dt$ above a threshold value f_0, where $\Omega(x)$ is the Heaviside step function ($\Omega(x) = 0$ if $x < 0$ and $\Omega(x) = 1$ otherwise). This feature reflects previous analysis of the same experimental apparatus [Johnson et al., 2013]. We use the thresholds of strain $f_0 = 10^{-9}$,5×10^{-9},10^{-8},5×10^{-8},10^{-7} (without unit).
- Time correlation features: power spectrum, autocorrelation, and partial autocorrelation. The power spectrum is given by $\|\hat{f}(\omega)\|^2$, with $\hat{f}(\omega) = \int f(t)e^{-2\pi i\omega t}\mathrm{d}t$ the Fourier transform of $f(t)$. To construct features, we integrate the power spectrum $\int_a^b \hat{f}(\omega)\mathrm{d}\omega$ using frequency windows $(a, b) = \{(19.65, 20.65), (39.8, 40.8), (80.1, 81.1)\}$ in kHz. The autocorrelation is $E[(X_t - \bar{X})(X_(t - h) - \bar{X})]/\mathrm{Var}(X_t)$, with $\bar{X}$ the mean of the series and timsecale $h = \frac{1}{41.25\mathrm{kHz}}$. The correlation runs over the values of t within the time window. The partial autocorrelation [Dodge, 2008] on the raw discretized time series f_t is given by $r(k) = \mathrm{Corr}(f_t, f_{t-k}|f_{t-1}, ..., f_{t-k+1})$,, with again the correlation running over the values of t within the time window and the discrete shift k corresponding to timescale $k = \frac{1}{41.25\mathrm{kHz}}$.

8.6 Conclusion

The results of our approach are highly suggestive. Does the laboratory tremor-like signal exist in brittle failure experiments and in granular flow such as in avalanche? Moreover, does it exist in faults in Earth? Since mature faults contain fault gouge, we may well expect that such signals are generated in Earth. If so, our results point toward a potentially rich signal that could be exploited for earthquake time forecasting as well as for forecasting in a broad spectrum of industrial and natural materials. Our work suggests that machine learning can be applied to acoustic and seismic data as a new means to reveal extremely informative, but as yet unidentified signals.

Chapter 9

Summary

In this thesis, we have studied and explored what simulation and the interplay between simulation and machine learning can bring to the understanding and improvement of optoelectronic devices, gallium nitride light emitting diodes in particular.

In the first part, we explored the use of Poisson-Schrödinger simulations for explaining and understanding experimental observations.

We have studied the effect of silicon doping on the emission properties of InGaN quantum wells, showing that simulations reproduce very precisely the experimentally observed shifts in emission wavelength, and explain them by the mitigation of the quantum confined Stark effect, mitigation due to the ionisation of the silicon in the quantum barriers.

We showed that simulations can explain experimentally observed emission inhomogeneities in the plane of InGaN quantum wells by the presence of p-doped material in the active region, brought there by hexagonal V-shaped defects.

In the second part we explored how machine learning can improve and leverage the use of optoelectronic device simulations.

We developed a machine learning algorithm that automatically adjusts the parameters of simulations to achieve convergence. Optoelectronic devices simulations have to tackle the resolution of very diverse equations at very different length scales, and as a result can suffer from poor convergence. Our algorithm finds the set of parameters that will achieve convergence, and can subsequently predict if a simulation will converge, thereby avoiding wasting time and computation resources on simulations that will not succeed.

Furthermore, we developed a machine learning algorithm that tunes the physical parameters of simulations to automatically reproduce experimen-

tal results. Poisson-Schrödinger simulations are very versatile, but rely on many physical parameters, owing to the different equations they solve self-consistently. As a result, their parameters must be tuned against experimental results before new simulation results can be trusted. Our algorithm is able to quickly and automatically find the set of physical parameters that reproduce an experiment, according to any given metric.

Finally, we developed a machine learning-based algorithm that automatically explores LED structures in order to maximise their efficiency. Coupled with the previous algorithms, it enables the discovery of new LED structures that have extremely high simulated efficiency, based on parameters previously validated on experimental results.

In the last part of the thesis, we have explored the use of machine learning for the analysis and prediction of material failure, in the setting of laboratory earthquakes. The method we developed is able to model and predict the upcoming failure of a sheet of compacted granular material using the acoustic signal it emits under shearing. In order to build its model and make its predictions, the algorithm was able to retrieve and isolate known acoustic precursors, but also discovered new signals that went unnoticed before, and may exist in the failure of other materials, including Earth faults.

Chapter 10

Future work

In this thesis, we have shown the value of machine learning in materials science for two situations: one where the underlying physics we are trying to uncover is very complicated and high-dimensional (optoelectronic device simulations), and the other where the amount of data is intractable (material failure).

Although machine learning is ubiquitous in disciplines that have a history for being data-driven, such as the medical sciences or some branches of computer science, it sometimes has a poor reputation in the physical sciences. Hopefully, the work showed here demonstrates that machine learning can be very valuable for the material scientist, as a way to efficiently build a very fast proxy for expensive models, and as a way to efficiently sift through large amounts of experimental data in an unbiased manner.

The work presented in this thesis had two main parts, and consequently has two main continuations.

The use of machine learning as an alternative to expensive physical models can be developed in a wide variety of situations, including multi-scale calculations where the smaller scale model is very computationally expensive, such as *ab-initio* molecular dynamics. The continuation of this first part of our work will be to develop other machine learning tools that leverage physical calculations by automatising and accelerating them.

The use of machine learning to parse vast amounts of experimental data can be used to uncover new physics. We showed that this is in part because the machine has no pre-conceived ideas and tries and tests all possible models. The continuation of this part of our work will be to apply the method to similar dataset such as the failure of other materials, including Earth faults.

Bibliography

[Abu-Mostafa et al., 2012] Abu-Mostafa, Y. S., Magdon-Ismail, M., and Lin, H.-T. (2012). *Learning from data*. AMLBook.

[Akasaka et al., 2004] Akasaka, T., Gotoh, H., Saito, T., and Makimoto, T. (2004). High luminescent efficiency of ingan multiple quantum wells grown on ingan underlying layers. *Appl. Phys. Lett.*, 85(15):3089–3091.

[Alam et al., 2000] Alam, M. A., Hybertsen, M. S., Smith, R. K., and Baraff, G. A. (2000). Simulation of semiconductor quantum well lasers. *IEEE Trans. Electron Devices*, 47(10):1917–1925.

[Azoff, 1989] Azoff, E. (1989). Energy transport numerical simulation of graded algaas/gaas heterojunction bipolar transistors. *Electron Devices, IEEE Transactions on*, 36(4):609–616.

[Badcock et al., 2013] Badcock, T. J., Hammersley, S., Watson-Parris, D., Dawson, P., Godfrey, M. J., Kappers, M. J., McAleese, C., Oliver, R. A., and Humphreys, C. J. (2013). Carrier density dependent localization and consequences for efficiency droop in InGaN/GaN quantum well structures. *Japanese Journal of Applied Physics*, 52:08JK10.

[Bakun and Lindh, 1985] Bakun, W. H. and Lindh, A. G. (1985). The Parkfield, California, Earthquake Prediction Experiment. *Science*, 229:619–624.

[Balachandran et al., 2015] Balachandran, P. V., Theiler, J., Rondinelli, J. M., and Lookman, T. (2015). Materials prediction via classification learning. *Sci. Rep.*, 5.

[Bergstra et al., 2011] Bergstra, J. S., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems*, pages 2546–2554.

[Bernardini and Fiorentini, 2000] Bernardini, F. and Fiorentini, V. (2000). Polarization fields in nitride nanostructures: 10 points to think about.

Appl. Surf. Sci., 166(14):23 – 29. 7TH International Conf. on Formation of Semiconductor Interfaces.

[Bernardini and Fiorentini, 2001] Bernardini, F. and Fiorentini, V. (2001). Nonlinear macroscopic polarization in iii-v nitride alloys. *Phys. Rev. B*, 64:085207.

[Bernardini et al., 1997] Bernardini, F., Fiorentini, V., and Vanderbilt, D. (1997). Spontaneous polarization and piezoelectric constants of iii-v nitrides. *Phys. Rev. B*, 56:R10024–R10027.

[Bertoldi and Atanasiu, 2007] Bertoldi, P. and Atanasiu, B. (2007). Electricity consumption and efficiency trends in the enlarged European Union. *IES–JRC. European Union.*

[Bhadeshia, 1999] Bhadeshia, H. K. D. H. (1999). Neural networks in materials science. *ISIJ Int.*, 39(10):966–979.

[Bogdanov et al., 2010] Bogdanov, M. V., Bulashevich, K. A., Khokhlev, O. V., Evstratov, I. Y., Ramm, M. S., and Karpov, S. Y. (2010). Current crowding effect on light extraction efficiency of thin-film leds. *physica status solidi (c)*, 7(7-8):2124–2126.

[Bouchon et al., 2013] Bouchon, M., Durand, V., Marsan, D., Karabulut, H., and Schmittbuhl, J. (2013). The long precursory phase of most large interplate earthquakes. *Nature Geosci*, 6(4):299–302.

[Bouchon et al., 2016] Bouchon, M., Marsan, D., Durand, V., Campillo, M., Perfettini, H., Madariaga, R., and Gardonio, B. (2016). Potential slab deformation and plunge prior to the tohoku, iquique and maule earthquakes. *Nature Geosci*, 9(5):380–383. Letter.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

[Burl and Wang, 2009] Burl, M. C. and Wang, E. (2009). Active learning for directed exploration of complex systems. *Proceedings of the 26th Annual International Conference on Machine Learning.*

[Calaprice, 2010] Calaprice, A. (2010). *The ultimate quotable Einstein.* Princeton University Press.

[Canali et al., 1975] Canali, C., Majni, G., Minder, R., and Ottaviani, G. (1975). Electron and hole drift velocity measurements in silicon and their

empirical relation to electric field and temperature. *Electron Devices, IEEE Transactions on*, 22(11):1045–1047.

[Castelli and Jacobsen, 2014] Castelli, I. E. and Jacobsen, K. W. (2014). Designing rules and probabilistic weighting for fast materials discovery in the perovskite structure. *Modell. Simul. Mater. Sci. Eng.*, 22(5).

[Cevik et al., 2016] Cevik, M., Ergun, M. A., Stout, N. K., Trentham-Dietz, A., Craven, M., and Alagoz, O. (2016). Using active learning for speeding up calibration in simulation models. *Med. Decis. Making*, 36(5):581–593.

[Chichibu et al., 2006] Chichibu, S. F., Uedono, A., Onuma, T., Haskell, B. A., Chakraborty, A., Koyama, T., Fini, P. T., Keller, S., Denbaars, S. P., Speck, J. S., Mishra, U. K., Nakamura, S., Yamaguchi, S., Kamiyama, S., Amano, H., Akasaki, I., Han, J., and Sota, T. (2006). Origin of defect-insensitive emission probability in in-containing (al, in, ga) n alloy semiconductors. *Nature Materials*, 5(10):810–816.

[Cho et al., 1998] Cho, Y.-H., Song, J., Keller, S., Minsky, M., Hu, E., Mishra, U., and DenBaars, S. (1998). Influence of si doping on characteristics of ingan/gan multiple quantum wells. *Appl. Phys. Lett.*, 73(8):1128–1130.

[Chouet, 1996] Chouet, B. A. (1996). Long-period volcano seismicity: its source and use in eruption forecasting. *Nature*, 380(6572):309–316.

[Chuang, 1991] Chuang, S. L. (1991). Efficient band-structure calculations of strained quantum wells. *Phys. Rev. B*, 43:9649–9661.

[Chuang and Chang, 1996] Chuang, S. L. and Chang, C. S. (1996). $\mathrm{k} \cdot \mathrm{p}$ method for strained wurtzite semiconductors. *Phys. Rev. B*, 54:2491–2504.

[Chuang and Chang, 1997] Chuang, S. L. and Chang, C. S. (1997). A band-structure model of strained quantum-well wurtzite semiconductors. *Semicond. Sci. Technol.*, 12(3):252.

[Dalfors et al., 1999] Dalfors, J., Bergman, J., Holtz, P., Sernelius, B., Monemar, B., Amano, H., and Akasaki, I. (1999). Optical properties of doped ingan/gan multiquantum-well structures. *Appl. Phys. Lett.*, 74(22):3299–3301.

[David et al., 2008] David, A., Grundmann, M. J., Kaeding, J. F., Gardner, N. F., Mihopoulos, T. G., and Krames, M. R. (2008). Carrier distribution in (0001)InGaN-GaN multiple quantum well light-emitting diodes. *Appl. Phys. Lett.*, 92(5).

[Davies et al., 2013] Davies, M. J., Badcock, T. J., Dawson, P., Kappers, M. J., Oliver, R. A., and Humphreys, C. J. (2013). High excitation carrier density recombination dynamics of InGaN/GaN quantum well structures: Possible relevance to efficiency droop. *Appl. Phys. Lett.*, 102(2):022106–022106–3.

[Davies et al., 2015] Davies, M. J., Dawson, P., Massabuau, F. C.-P., Fol, A. L., Oliver, R. A., Kappers, M. J., and Humphreys, C. J. (2015). A study of the inclusion of prelayers in ingan/gan single-and multiple-quantum-well structures. *physica status solidi (b)*, 252(5):866–872.

[Deguchi et al., 1998] Deguchi, T., Shikanai, A., Torii, K., Sota, T., Chichibu, S., and Nakamura, S. (1998). Luminescence spectra from ingan multiquantum wells heavily doped with si. *Appl. Phys. Lett.*, 72(25):3329–3331.

[Delaney et al., 2009] Delaney, K. T., Rinke, P., and Van de Walle, C. G. (2009). Auger recombination rates in nitrides from first principles. *Appl. Phys. Lett.*, 94(19):191109–.

[Delorey et al., 2017] Delorey, A. A., van der Elst, N. J., and Johnson, P. A. (2017). Tidal triggering of earthquakes suggests poroelastic behavior on the san andreas fault. *Earth Planet. Sci. Lett.*, 460:164–170.

[Dodge, 2008] Dodge, Y. (2008). *The concise encyclopedia of statistics*. Springer Science & Business Media.

[DOE, 2010a] DOE (2010a). *Buildings Energy Data Book, Chapter 1.1 Building Sector Energy Consumption*. United States Department of Energy.

[DOE, 2010b] DOE (2010b). *Buildings Energy Data Book, Chapter 2.1 Residential Sector Energy Consumption*. United States Department of Energy.

[DOE, 2010c] DOE (2010c). *Buildings Energy Data Book, Chapter 3.1 Commercial Sector Energy Consumption*. United States Department of Energy.

[Domingos, 2000] Domingos, P. (2000). A unified bias-variance decomposition. In *Proceedings of 17th International Conference on Machine Learning. Stanford CA Morgan Kaufmann*, pages 231–238.

[Domingos, 2012] Domingos, P. (2012). A few useful things to know about machine learning. *Commun. ACM*, 55(10):78–87.

[Efremov et al., 2006] Efremov, A., Bochkareva, N., Gorbunov, R., Lavrinovich, D., Rebane, Y., Tarkhin, D., and Shreter, Y. (2006). Effect of the joule heating on the quantum efficiency and choice of thermal conditions for high-power blue InGaN/GaN LEDs. *Semiconductors*, 40(5):605–610.

[Eiting et al., 1998] Eiting, C. J., Grudowski, P. A., Dupuis, R. D., Hsia, H., Tang, Z., Becher, D., Kuo, H., Stillman, G. E., and Feng, M. (1998). Activation studies of low-dose si implants in gallium nitride. *Appl. Phys. Lett.*, 73(26):3875–3877.

[Farrell and Gartland Jr, 1991] Farrell, P. A. and Gartland Jr, E. C. (1991). *On the Scharfetter-Gummel discretization for drift-diffusion continuity equations*, volume 1. Boole Dublin.

[Ferdowsi et al., 2014] Ferdowsi, B., Griffa, M., Guyer, R. A., Johnson, P. A., Marone, C., and Carmeliet, J. (2014). Three-dimensional discrete element modeling of triggered slip in sheared granular media. *Phys. Rev. E*, 89:042204.

[Ferdowsi et al., 2015] Ferdowsi, B., Griffa, M., Guyer, R. A., Johnson, P. A., Marone, C., and Carmeliet, J. (2015). Acoustically-induced slip in sheared granular layers: application to dynamic earthquake triggering. *Geophys. Res. Lett.*, 42(22):9750–9757.

[Fiorentini et al., 2002] Fiorentini, V., Bernardini, F., and Ambacher, O. (2002). Evidence for nonlinear macroscopic polarization in iii-v nitride alloy heterostructures. *Appl. Phys. Lett.*, 80(7):1204–1206.

[Fiorentini et al., 1999] Fiorentini, V., Bernardini, F., Della Sala, F., Di Carlo, A., and Lugli, P. (1999). Effects of macroscopic polarization in iii-v nitride multiple quantum wells. *Phys. Rev. B*, 60(12):8849.

[Frazier et al., 2009] Frazier, P., Powell, W., and Dayanik, S. (2009). The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21(4):599–613.

[Gardner et al., 2007] Gardner, N. F., Muller, G. O., Shen, Y. C., Chen, G., Watanabe, S., Gotz, W., and Krames, M. R. (2007). Blue-emitting InGaN-GaN double-heterostructure light-emitting diodes reaching maximum quantum efficiency above 200A/cm2. *Appl. Phys. Lett.*, 91(24):243506.

[Geller et al., 1997] Geller, R. J., Jackson, D. D., Kagan, Y. Y., and Mulargia, F. (1997). Earthquakes cannot be predicted. *Science*, 275(5306):1616–1617.

[Ghiringhelli et al., 2015] Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C., and Scheffler, M. (2015). Big data of materials science: Critical role of the descriptor. *Phys. Rev. Lett.*, 114(10).

[Goebel et al., 2013] Goebel, T., Schorlemmer, D., Becker, T., Dresen, G., and Sammis, C. (2013). Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments. *Geophys. Res. Lett.*, 40(10):2049–2054.

[Goldfinger et al., 2016] Goldfinger, C., Galer, S., Beeson, J., Hamilton, T., Black, B., Romsos, C., Patton, J., Nelson, C. H., Hausmann, R., and Morey, A. (2016). The importance of site selection, sediment supply, and hydrodynamics: A case study of submarine paleoseismology on the northern cascadia margin, washington usa. *Mar. Geol.*

[Gramacy et al., 2004] Gramacy, R. B., Lee, H. K. H., and Macready, W. G. (2004). Parameter space exploration with gaussian process trees. In *Proceedings of the Twenty-first International Conference on Machine Learning*, ICML '04, pages 45–, New York, NY, USA. ACM.

[Griffiths et al., 2015] Griffiths, J. T., Zhang, S., Rouet-Leduc, B., Fu, W. Y., Bao, A., Zhu, D., Wallis, D. J., Howkins, A., Boyd, I., Stowe, D., et al. (2015). Nanocathodoluminescence reveals mitigation of the stark shift in InGaN quantum wells by Si doping. *Nano Lett.*, 15(11):7639–7643.

[Grinberg et al., 1984] Grinberg, A. A., Shur, M., Fischer, R., and Morkoc, H. (1984). An investigation of the effect of graded layers and tunneling on the performance of AlGaAs/GaAs heterojunction bipolar transistors. *Electron Devices, IEEE Transactions on*, 31(12):1758–1765.

[Gupta and Miescke, 1996] Gupta, S. S. and Miescke, K. J. (1996). Bayesian look ahead one-stage sampling allocations for selection of the best population. *Journal of Statistical Planning and Inference*, 54(2):229–244. 40 Years of Statistical Selection Theory, Part I.

[Gutenberg and Richter, 1949] Gutenberg, B. and Richter, C. (1949). *Seismicity of the Earth, 273.* Princeton University Press, Princeton, NJ.

[Hader et al., 2008] Hader, J., Moloney, J. V., Pasenow, B., Koch, S. W., Sabathil, M., Linder, N., and Lutgen, S. (2008). On the importance of

radiative and auger losses in GaN-based quantum wells. *Appl. Phys. Lett.*, 92(26):–.

[Hager et al., 2009] Hager, C. E., Jones, K. A., Derenge, M. A., and Zheleva, T. S. (2009). Activation of ion implanted si in gan using a dual aln annealing cap. *J. Appl. Phys.*, 105(3):–.

[Haitz and Tsao, 2011] Haitz, R. and Tsao, J. Y. (2011). Solid-state lighting: 'the case' 10 years after and future prospects. *physica status solidi (a)*, 208(1):17–29.

[Hammersley et al., 2012] Hammersley, S., Watson-Parris, D., Dawson, P., Godfrey, M. J., Badcock, T. J., Kappers, M. J., McAleese, C., Oliver, R. A., and Humphreys, C. J. (2012). The consequences of high injected carrier densities on carrier localization and efficiency droop in InGaN/GaN quantum well structures. *J. Appl. Phys.*, 111(8):–.

[Han et al., 2009] Han, S.-H., Lee, D.-Y., Lee, S.-J., Cho, C.-Y., Kwon, M.-K., Lee, S. P., Noh, D. Y., Kim, D.-J., Kim, Y. C., and Park, S.-J. (2009). Effect of electron blocking layer on efficiency droop in InGaN/GaN multiple quantum well light-emitting diodes. *Appl. Phys. Lett.*, 94(23):–.

[Han et al., 2013] Han, S.-H., Lee, D.-Y., Shim, H.-W., Lee, J. W., Kim, D.-J., Yoon, S., Kim, Y. S., and Kim, S.-T. (2013). Improvement of efficiency and electrical properties using intentionally formed v-shaped pits in ingan/gan multiple quantum well light-emitting diodes. *Appl. Phys. Lett.*, 102(25):251123.

[Hangleiter et al., 2005] Hangleiter, A., Hitzel, F., Netzel, C., Fuhrmann, D., Rossow, U., Ade, G., and Hinze, P. (2005). Suppression of nonradiative recombination by v-shaped pits in GaInN/GaN quantum wells produces a large increase in the light emission efficiency. *Phys. Rev. Lett.*, 95:127402.

[Hangleiter et al., 2007] Hangleiter, A., Netzel, C., Fuhrmann, D., Hitzel, F., Hoffmann, L., Bremers, H., Rossow, U., Ade, G., and Hinze, P. (2007). Anti-localization suppresses non-radiative recombination in GaInN/GaN quantum wells. *Philos. Mag.*, 87(13):2041–2065.

[Hautier et al., 2010] Hautier, G., Fischer, C. C., Jain, A., Mueller, T., and Ceder, G. (2010). Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chem. Mater.*, 22(12):3762–3767.

[Ho, 1998] Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(8):832–844.

[Holonyak and Bevacqua, 1962] Holonyak, N. and Bevacqua, S. F. (1962). Coherent (visible) light emission from ga(as1-xpx) junctions. *Appl. Phys. Lett.*, 1(4):82–83.

[Huang et al., 1998] Huang, M., Jiang, L., Liaw, P. K., Brooks, C. R., Seeley, R., and Klarstrom, D. L. (1998). Using acoustic emission in fatigue and fracture materials research. *JOM*, 50(11):1–14.

[Humphreys et al., 2017] Humphreys, C., Griffiths, J., Tang, F., Oehler, F., Findlay, S., Zheng, C., Etheridge, J., Martin, T., Bagot, P., Moody, M., Sutherland, D., Dawson, P., Schulz, S., Zhang, S., Fu, W., Zhu, T., Kappers, M., and Oliver, R. (2017). The atomic structure of polar and non-polar ingan quantum wells and the green gap problem. *Ultramicroscopy*, 176:93 – 98. 70th Birthday of Robert Sinclair and 65th Birthday of Nestor J. Zaluzec {PICO} 2017 âĂŞ Fourth Conference on Frontiers of Aberration Corrected Electron Microscopy.

[Humphreys, 2008] Humphreys, C. J. (2008). Solid-state lighting. *MRS Bull.*, 33:459–470.

[Hutter et al., 2011] Hutter, F., Hoos, H. H., and Leyton-Brown, K. (2011). *Sequential Model-Based Optimization for General Algorithm Configuration*, pages 507–523. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Irokawa et al., 2005] Irokawa, Y., Fujishima, O., Kachi, T., and Nakano, Y. (2005). Electrical activation characteristics of silicon-implanted gan. *J. Appl. Phys.*, 97(8):–.

[Iveland et al., 2013] Iveland, J., Martinelli, L., Peretti, J., Speck, J. S., and Weisbuch, C. (2013). Direct measurement of Auger electrons emitted from a semiconductor light-emitting diode under electrical injection: Identification of the dominant mechanism for efficiency droop. *Phys. Rev. Lett.*, 110:177406.

[Jaeger et al., 2009] Jaeger, J. C., Cook, N. G., and Zimmerman, R. (2009). *Fundamentals of rock mechanics*. John Wiley & Sons.

[Jain et al., 2013] Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., and

Persson, K. A. (2013). Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.*, 1(1):011002.

[Johnson et al., 2016] Johnson, P. A., Carmeliet, J., Savage, H. M., Scuderi, M., Carpenter, B. M., Guyer, R. A., Daub, E. G., and Marone, C. (2016). Dynamically triggered slip leading to sustained fault gouge weakening under laboratory shear conditions. *Geophys. Res. Lett.*, 43(4):1559–1565. 2015GL067056.

[Johnson et al., 2013] Johnson, P. A., Ferdowsi, B., Kaproth, B. M., Scuderi, M., Griffa, M., Carmeliet, J., Guyer, R. A., Le Bas, P.-Y., Trugman, D. T., and Marone, C. (2013). Acoustic emission and microslip precursors to stick-slip failure in sheared granular material. *Geophys. Res. Lett.*, 40(21):5627–5631. 2013GL057848.

[Jones et al., 1998] Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *J. Global Optim.*, 13(4):455–492.

[Jordan et al., 2011] Jordan, T., Chen, Y.-T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., Papadopoulos, G., Yamaoka, K., and Zschau, J. (2011). Operational earthquake forecasting: State of knowledge and guidelines for implementation. *Annals of Geophysics.*

[Journel and Huijbregts, 1978] Journel, A. and Huijbregts, C. J. (1978). *Mining Geostatistics.* Academic Press, London.

[Kang et al., 2015] Kang, D., Kim, T., Song, K., Back, J., Jeong, H., Song, J.-O., and Seong, T.-Y. (2015). Improving output power performance of InGaN-based light-emitting diodes by employing step-down indium contents. *Japanese Journal of Applied Physics*, 54(4):042102.

[Kim et al., 2015] Kim, D. Y., Lin, G.-B., Hwang, S., Park, J. H., Meyaard, D., Schubert, E., Ryu, H.-Y., and Kim, J. K. (2015). Polarization-engineered high-efficiency gainn light-emitting diodes optimized by genetic algorithm. *Photonics Journal, IEEE*, 7(1):1–9.

[Kim et al., 2007] Kim, M.-H., Schubert, M. F., Dai, Q., Kim, J. K., Schubert, E. F., Piprek, J., and Park, Y. (2007). Origin of efficiency droop in GaN-based light-emitting diodes. *Appl. Phys. Lett.*, 91(18):–.

[Kioupakis et al., 2011] Kioupakis, E., Rinke, P., Delaney, K., and Van de Walle, C. G. (2011). Indirect auger recombination as a cause of efficiency

droop in nitride light-emitting diodes. *Appl. Phys. Lett.*, 98(16):161107–161107–3.

[Kleinberg, 1996] Kleinberg, E. (1996). An overtraining-resistant stochastic modeling method for pattern recognition. *The annals of statistics*, 24(6):2319–2349.

[Laubsch et al., 2009] Laubsch, A., Sabathil, M., Bergbauer, W., Strassburg, M., Lugauer, H., Peter, M., Lutgen, S., Linder, N., Streubel, K., Hader, J., Moloney, J. V., Pasenow, B., and Koch, S. W. (2009). On the origin of IQE-droop in InGaN LEDs. *physica status solidi (c)*, 6(S2):S913–S916.

[Li et al., 2012] Li, X., Okur, S., Zhang, F., Avrutin, V., Ozgur, U., Morkoc, H., Hong, S. M., Yen, S., Hsu, T., and Matulionis, A. (2012). Impact of active layer design on InGaN radiative recombination coefficient and LED performance. *J. Appl. Phys.*, 111(6):063112.

[Lim et al., 2009] Lim, S. K., Brewster, M., Qian, F., Li, Y., Lieber, C. M., and Gradecak, S. (2009). Direct correlation between structural and optical properties of iii- v nitride nanowire heterostructures with nanoscale resolution. *Nano Lett.*, 9(11):3940–3944.

[Lin et al., 2012a] Lin, G.-B., Meyaard, D., Cho, J., Fred Schubert, E., Shim, H., and Sone, C. (2012a). Analytic model for the efficiency droop in semiconductors with asymmetric carrier-transport properties based on drift-induced reduction of injection efficiency. *Appl. Phys. Lett.*, 100(16):161106–161106–4.

[Lin et al., 2012b] Lin, Y., Zhang, Y., Liu, Z., Su, L., Zhang, J., Wei, T., and Chen, Z. (2012b). Spatially resolved study of quantum efficiency droop in ingan light-emitting diodes. *Appl. Phys. Lett.*, 101(25):252103.

[Liu et al., 2008a] Liu, J. P., Limb, J. B., Ryou, J.-H., Yoo, D., Horne, C. A., Dupuis, R. D., Wu, Z. H., Fischer, A. M., Ponce, F. A., Hanser, A. D., Liu, L., Preble, E. A., and Evans, K. R. (2008a). Blue light emitting diodes grown on freestanding (11-20) a-plane gan substrates. *Appl. Phys. Lett.*, 92(1):–.

[Liu et al., 2008b] Liu, J. P., Ryou, J.-H., Dupuis, R. D., Han, J., Shen, G. D., and Wang, H. B. (2008b). Barrier effect on hole transport and carrier distribution in InGaN/GaN multiple quantum well visible light-emitting diodes. *Appl. Phys. Lett.*, 93(2):–.

[Longini and Greene, 1956] Longini, R. L. and Greene, R. F. (1956). Ionization interaction between impurities in semiconductors and insulators. *Phys. Rev.*, 102:992–999.

[Lophaven et al., 2002] Lophaven, S. N., Nielsen, H. B., and Søndergaard, J. (2002). Dace-a matlab kriging toolbox, version 2.0. Technical report.

[Love, 1944] Love, A. (1944). *A Treatise on the Mathematical theory of Elasticity*. New York, Dover Publication.

[Luttinger and Kohn, 1955] Luttinger, J. M. and Kohn, W. (1955). Motion of electrons and holes in perturbed periodic fields. *Phys. Rev.*, 97:869–883.

[Manyika et al., 2011] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. (2011). Big data: The next frontier for innovation, competition. Technical report, and productivity. Technical report, McKinsey Global Institute.

[Marone, 1998] Marone, C. (1998). Laboratory-derived friction laws and their application to seismic faulting. *Annu. Rev. Earth Planet. Sci.*, 26(1):643–696.

[McGuire et al., 2015] McGuire, J. J., Lohman, R. B., Catchings, R. D., Rymer, M. J., and Goldman, M. R. (2015). Relationships among seismic velocity, metamorphism, and seismic and aseismic fault slip in the salton sea geothermal field region. *Journal of Geophysical Research: Solid Earth*, 120(4):2600–2615. 2014JB011579.

[Melbourne and Webb, 2003] Melbourne, T. I. and Webb, F. H. (2003). Slow but not quite silent. *Science*, 300(5627):1886–1887.

[Merano et al., 2005] Merano, M., Sonderegger, S., Crottini, A., Collin, S., Renucci, P., Pelucchi, E., Malko, A., Baier, M., Kapon, E., Deveaud, B., et al. (2005). Probing carrier dynamics in nanostructures by picosecond cathodoluminescence. *Nature*, 438(7067):479–482.

[Michlmayr et al., 2013] Michlmayr, G., Cohen, D., and Or, D. (2013). Shear-induced force fluctuations and acoustic emissions in granular material. *Journal of Geophysical Research: Solid Earth*, 118(12):6086–6098.

[Mignan, 2014] Mignan, A. (2014). The debate on the prognostic value of earthquake foreshocks: A meta-analysis. *Sci. Rep.*, 4:4099.

[Miller et al., 1984] Miller, D. A. B., Chemla, D. S., Damen, T. C., Gossard, A. C., Wiegmann, W., Wood, T. H., and Burrus, C. A. (1984). Band-edge electroabsorption in quantum well structures: The quantum-confined stark effect. *Phys. Rev. Lett.*, 53:2173–2176.

[Mitchell, 1997] Mitchell, T. M. (1997). Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, 45.

[Naik et al., 2013] Naik, A. W., Kangas, J. D., Langmead, C. J., and Murphy, R. F. (2013). Efficient modeling and active learning discovery of biological responses. *PLoS One*, 8(12):e83996.

[Naik et al., 2016] Naik, A. W., Kangas, J. D., Sullivan, D. P., and Murphy, R. F. (2016). Active machine learning-driven experimentation to determine compound effects on protein patterns. *eLife*, 5:e10047.

[Nakamura et al., 1995] Nakamura, S., Senoh, M., Iwasa, N., ichi Nagahama, S., Yamada, T., and Mukai, T. (1995). Superbright green ingan single-quantum-well-structure light-emitting diodes. *Japanese Journal of Applied Physics*, 34(10B):L1332.

[Nakamura et al., 1993] Nakamura, S., Senoh, M., and Mukai, T. (1993). P-gan/n-ingan/n-gan double-heterostructure blue-light-emitting diodes. *Japanese Journal of Applied Physics Part 2 Letters*, 32:L8.

[Ni et al., 2008] Ni, X., Fan, Q., Shimada, R., Ozgur, U., and Morkoc, H. (2008). Reduction of efficiency droop in InGaN light emitting diodes by coupled quantum wells. *Appl. Phys. Lett.*, 93(17):–.

[Niemeijer et al., 2010] Niemeijer, A., Marone, C., and Elsworth, D. (2010). Frictional strength and strain weakening in simulated fault gouge: Competition between geometrical weakening and chemical strengthening. *Journal of Geophysical Research: Solid Earth*, 115(B10).

[Obara, 2002] Obara, K. (2002). Nonvolcanic deep tremor associated with subduction in southwest japan. *Science*, 296(5573):1679–1681.

[O'Hagan and Kingman, 1978] O'Hagan, A. and Kingman, J. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–42.

[Oliver et al., 2010] Oliver, R. A., Bennett, S. E., Zhu, T., Beesley, D. J., Kappers, M. J., Saxey, D. W., Cerezo, A., and Humphreys, C. J. (2010). Microstructural origins of localization in InGaN quantum wells. *J. Phys. D: Appl. Phys.*, 43(35):354003.

[Oliver et al., 2013] Oliver, R. A., Massabuau, F. C.-P., Kappers, M. J., Phillips, W. A., Thrush, E. J., Tartan, C. C., Blenkhorn, W. E., Badcock, T. J., Dawson, P., Hopkins, M. A., Allsopp, D. W. E., and Humphreys, C. J. (2013). The impact of gross well width fluctuations on the efficiency of gan-based light emitting diodes. *Appl. Phys. Lett.*, 103(14):141114.

[Olshansky et al., 1984] Olshansky, R., Su, C., Manning, J., and Powazinik, W. (1984). Measurement of radiative and nonradiative recombination rates in InGaAsp and AlGaAs light sources. *Quantum Electronics, IEEE Journal of*, 20(8):838–854.

[Osborne et al., 2009] Osborne, M. A., Garnett, R., and Roberts, S. J. (2009). Gaussian processes for global optimization. In *3rd international conference on learning and intelligent optimization (LION3)*, pages 1–15.

[Otsuji et al., 2006] Otsuji, N., Fujiwara, K., and Sheu, J.-K. (2006). Electroluminescence efficiency of blue in ga n/ ga n quantum-well diodes with and without an n-in ga n electron reservoir layer. *J. Appl. Phys.*, 100(11):113105.

[Pantzas et al., 2015] Pantzas, K., Patriarche, G., Troadec, D., Kociak, M., Cherkashin, N., Hÿtch, M., Barjon, J., Tanguy, C., Rivera, T., Suresh, S., et al. (2015). Role of compositional fluctuations and their suppression on the strain and luminescence of ingan alloys. *J. Appl. Phys.*, 117(5):055705.

[Park et al., 2007] Park, E.-H., Kang, D. N. H., Ferguson, I. T., Park, S.-K. J. J.-S., and Yoo, T.-K. (2007). The effect of silicon doping in the selected barrier on the electroluminescence of in ga n/ ga n multiquantum well light emitting diode. *Appl. Phys. Lett.*, 90(3):031102.

[Park, 2002] Park, S.-H. (2002). Crystal orientation effects on electronic properties of wurtzite ingan/gan quantum wells. *J. Appl. Phys.*, 91(12):9904–9908.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12:2825–2830.

[Petrov and Gvozdover, 1991] Petrov, V. and Gvozdover, R. (1991). Spatial resolution of cathodoluminescence scanning electron microscopy of semiconductors. *Scanning*, 13(6):410–414.

[Pfingsten, 2006] Pfingsten, T. (2006). Bayesian active learning for sensitivity analysis. In *ECML 2006*, pages 353–364, Berlin, Germany. Max-Planck-Gesellschaft, Springer.

[Pilania et al., 2013] Pilania, G., Wang, C., Jiang, X., Rajasekaran, S., and Ramprasad, R. (2013). Accelerating materials property predictions using machine learning. *Sci. Rep.*, 3.

[Piprek, 2007] Piprek, J. (2007). *Nitride semiconductor devices: principles and simulation*. John Wiley & Sons.

[Piprek, 2010] Piprek, J. (2010). Efficiency droop in nitride-based light-emitting diodes. *physica status solidi (a)*, 207(10):2217–2225.

[Piprek, 2014] Piprek, J. (2014). Origin of ingan/gan light-emitting diode efficiency improvements using tunnel-junction-cascaded active regions. *Appl. Phys. Lett.*, 104(5):–.

[Pomerleau, 1989] Pomerleau, D. A. (1989). Alvinn: An autonomous land vehicle in a neural network. Technical report, DTIC Document.

[Pope et al., 2003] Pope, I. A., Smowton, P. M., Blood, P., Thomson, J. D., Kappers, M. J., and Humphreys, C. J. (2003). Carrier leakage in InGaN quantum well light-emitting diodes emitting at 480 nm. *Appl. Phys. Lett.*, 82(17):2755–2757.

[Pradhan et al., 2006] Pradhan, S., Hansen, A., and Hemmer, P. C. (2006). Crossover behavior in failure avalanches. *Phys. Rev. E*, 74:016122.

[Rajan, 2005] Rajan, K. (2005). Materials informatics. *Mater. Today*, 8(10):38–45.

[Rasmussen, 2006] Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. MIT Press.

[Ren et al., 2016] Ren, C., Rouet-Leduc, B., Griffiths, J., Bohacek, E., Wallace, M., Edwards, P., Hopkins, M., Allsopp, D., Kappers, M., Martin, R., et al. (2016). Analysis of defect-related inhomogeneous electroluminescence in InGaN/GaN QW LEDs. *Superlattices Microstruct.*, 99:118–124.

[Ren, 2015] Ren, C. X. (2015). Polarisation fields in III-nitrides: effects and control. *Mater. Sci. Technol.*, 0(0):1–16.

[Rogers and Dragert, 2003] Rogers, G. and Dragert, H. (2003). Episodic tremor and slip on the cascadia subduction zone: The chatter of silent slip. *Science*, 300(5627):1942–1943.

[Rouet-Leduc et al., 2016] Rouet-Leduc, B., Barros, K., Lookman, T., and Humphreys, C. J. (2016). Optimisation of GaN LEDs and the reduction of efficiency droop using active machine learning. *Sci. Rep.*, 6:24862.

[Rouet-Leduc et al., 2017] Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C., and Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *arXiv preprint arXiv:1702.05774*.

[Rozhansky and Zakheim, 2007] Rozhansky, I. V. and Zakheim, D. A. (2007). Analysis of processes limiting quantum efficiency of algainn leds at high pumping. *physica status solidi (a)*, 204(1):227–230.

[Rubinstein et al., 2009] Rubinstein, J. L., Shelly, D. R., and Ellsworth, W. L. (2009). Non-volcanic tremor: A window into the roots of fault zones. In *New Frontiers in Integrated Solid Earth Sciences*, pages 287–314. Springer.

[Ryou et al., 2008] Ryou, J.-H., Limb, J., Lee, W., Liu, J., Lochner, Z., Yoo, D., and Dupuis, R. D. (2008). Effect of silicon doping in the quantum-well barriers on the electrical and optical properties of visible green light-emitting diodes. *Photonics Technology Letters, IEEE*, 20(21):1769–1771.

[Ryou et al., 2009] Ryou, J.-H., Yoder, P., Liu, J., Lochner, Z., Kim, H., Choi, S., Kim, H.-J., and Dupuis, R. D. (2009). Control of quantum-confined stark effect in ingan-based quantum wells. *Selected Topics in Quantum Electronics, IEEE Journal of*, 15(4):1080–1091.

[Sacks et al., 1989] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). *Design and analysis of computer experiments*.

[Sarkissian et al., 2013] Sarkissian, R., Roberts, S. T., Yeh, T.-W., Das, S., Bradforth, S. E., O'Brien, J., and Daniel Dapkus, P. (2013). Photon quenching in InGaN quantum well light emitting devices. *Appl. Phys. Lett.*, 103(4):–.

[Scharfetter and Gummel, 1969] Scharfetter, D. and Gummel, H. (1969). Large-signal analysis of a silicon read diode oscillator. *Electron Devices, IEEE Transactions on*, 16(1):64–77.

[Schmidt et al., 2012] Schmidt, G., Müller, M., Bertram, F., Veit, P., Petzold, S., Das, A., and Monroy, E. (2012). Highly spatially resolved cathodoluminescence of single gan quantum dots directly performed in a scanning transmission electron microscope. *Microsc. Microanal.*, 18(S2):1878–1879.

[Scholz et al., 1973] Scholz, C. H., Sykes, L. R., and Aggarwal, Y. P. (1973). Earthquake prediction: A physical basis. *Science*, 181(4102):803–810.

[Schomig et al., 2004] Schomig, H., Halm, S., Forchel, A., Bacher, G., Off, J., and Scholz, F. (2004). Probing individual localization centers in an InGaN/GaN quantum well. *Phys. Rev. Lett.*, 92:106802.

[Schubert et al., 2005] Schubert, E. F., Gessmann, T., and Kim, J. K. (2005). *Light emitting diodes*. Wiley Online Library.

[Schubert et al., 2007] Schubert, M. F., Chhajed, S., Kim, J. K., Schubert, E. F., Koleske, D. D., Crawford, M. H., Lee, S. R., Fischer, A. J., Thaler, G., and Banas, M. A. (2007). Effect of dislocation density on efficiency droop in GaInN/GaN light-emitting diodes. *Appl. Phys. Lett.*, 91(23):–.

[Schubert et al., 2009] Schubert, M. F., Xu, J., Dai, Q., Mont, F. W., Kim, J. K., and Schubert, E. F. (2009). On resonant optical excitation and carrier escape in GaInN/GaN quantum wells. *Appl. Phys. Lett.*, 94(8):081114.

[Schubert et al., 2008] Schubert, M. F., Xu, J., Kim, J. K., Schubert, E. F., Kim, M. H., Yoon, S., Lee, S. M., Sone, C., Sakong, T., and Park, Y. (2008). Polarization-matched GaInN-AlGaInN multi-quantum-well light-emitting diodes with reduced efficiency droop. *Appl. Phys. Lett.*, 93(4):–.

[Schubnel et al., 2013] Schubnel, A., Brunet, F., Hilairet, N., Gasc, J., Wang, Y., and Green, H. W. (2013). Deep-focus earthquake analogs recorded at high pressure and temperature in the laboratory. *Science*, 341(6152):1377–1380.

[Schulz et al., 2015] Schulz, S., Caro, M. A., Coughlan, C., and O'Reilly, E. P. (2015). Atomistic analysis of the impact of alloy and well-width fluctuations on the electronic and optical properties of ingan/gan quantum wells. *Phys. Rev. B*, 91:035439.

[Schwach et al., 2007] Schwach, C., Weisbuch, C., DenBaars, S., Benisty, H., and Nakamura, S. (2007). White, single or multi-color light emitting diodes by recycling guided modes. US Patent 7,223,998.

[Schwartz and Coppersmith, 1984] Schwartz, D. P. and Coppersmith, K. J. (1984). Fault behavior and characteristic earthquakes: Examples from the wasatch and san andreas fault zones. *Journal of Geophysical Research: Solid Earth*, 89(B7):5681–5698.

[Scornet, 2016] Scornet, E. (2016). On the asymptotics of random forests. *Journal of Multivariate Analysis*, 146:72–83. Special Issue on Statistical Models and Methods for High or Infinite Dimensional Spaces.

[Scuderi et al., 2014] Scuderi, M. M., Carpenter, B. M., and Marone, C. (2014). Physicochemical processes of frictional healing: Effects of water on stick-slip stress drop and friction of granular fault gouge. *Journal of Geophysical Research: Solid Earth*, 119(5):4090–4105.

[Seo Im et al., 1998] Seo Im, J., Kollmer, H., Off, J., Sohmer, A., Scholz, F., and Hangleiter, A. (1998). Reduction of oscillator strength due to piezoelectric fields in gan/alxga1-xn quantum wells. *Phys. Rev. B*, 57:R9435–R9438.

[Shan et al., 2011] Shan, Q., Meyaard, D. S., Dai, Q., Cho, J., Schubert, E. F., Son, J. K., and Sone, C. (2011). Transport-mechanism analysis of the reverse leakage current in gainn light-emitting diodes. *Appl. Phys. Lett.*, 99(25):253506.

[Shelly et al., 2007] Shelly, D. R., Beroza, G. C., and Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, 446(7133):305–307.

[Shen et al., 2007] Shen, Y. C., Mueller, G. O., Watanabe, S., Gardner, N. F., Munkholm, A., and Krames, M. R. (2007). Auger recombination in InGaN measured by photoluminescence. *Appl. Phys. Lett.*, 91(14):–.

[Sheu et al., 2003] Sheu, J. K., Chang, S., Kuo, C.-H., Su, Y., Wu, L., Lin, Y., Lai, W., Tsai, J. M., Chi, G., and Wu, R. K. (2003). White-light emission from near uv ingan-gan led chip precoated with blue/green/red phosphors. *Photonics Technology Letters, IEEE*, 15(1):18–20.

[Shockley and Read, 1952] Shockley, W. and Read, W. T. (1952). Statistics of the recombinations of holes and electrons. *Phys. Rev.*, 87:835–842.

[Smith and Smith, 1978] Smith, R. A. and Smith, R. A. (1978). *Semiconductors*, volume 173. Cambridge University Press Cambridge.

[Sonderegger et al., 2006] Sonderegger, S., Feltin, E., Merano, M., Crottini, A., Carlin, J., Sachot, R., Deveaud, B., Grandjean, N., and Ganiere, J. (2006). High spatial resolution picosecond cathodoluminescence of ingan quantum wells. *Appl. Phys. Lett.*, 89(23):232109.

[Srinivas et al., 2012] Srinivas, N., Krause, A., Kakade, S., and Seeger, M. (2012). Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *Information Theory, IEEE Transactions on*, 58(5):3250–3265.

[Stanton et al., 2001] Stanton, N. M., Kent, A. J., Akimov, A. V., Hawker, P., Cheng, T. S., and Foxon, C. T. (2001). Energy relaxation by hot electrons in n-gan epilayers. *J. Appl. Phys.*, 89(2):973–979.

[Stevenson, 2009] Stevenson, R. (2009). The LED's dark secret. *Spectrum, IEEE*, 46(8):26–31.

[Stratton, 1972] Stratton, R. (1972). Semiconductor current-flow equations (diffusion and degeneracy). *Electron Devices, IEEE Transactions on*, 19(12):1288–1292.

[Sutton and Barto, 1998] Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge.

[Suzuki et al., 1995] Suzuki, M., Uenoyama, T., and Yanase, A. (1995). First-principles calculations of effective-mass parameters of aln and gan. *Phys. Rev. B*, 52:8132–8139.

[Sze and Ng, 2006] Sze, S. M. and Ng, K. K. (2006). *Physics of semiconductor devices*. John Wiley & Sons.

[Takahashi et al., 2004] Takahashi, Y., Satake, A., Fujiwara, K., Shue, J., Jahn, U., Kostial, H., and Grahn, H. (2004). Enhanced radiative efficiency in blue (in, ga) n multiple-quantum-well light-emitting diodes with an electron reservoir layer. *Physica E*, 21(2):876–880.

[Takeuchi et al., 1997] Takeuchi, T., Sota, S., Katsuragawa, M., Komori, M., Takeuchi, H., Amano, H., and Akasaki, I. (1997). Quantum-confined stark effect due to piezoelectric fields in gainn strained quantum wells. *Japanese Journal of Applied Physics*, 36(4A):L382.

[Teo et al., 1998] Teo, K. L., Colton, J. S., Yu, P. Y., Weber, E. R., Li, M. F., Liu, W., Uchida, K., Tokunaga, H., Akutsu, N., and Matsumoto, K. (1998). An analysis of temperature dependent photoluminescence line shapes in InGaN. *Appl. Phys. Lett.*, 73(12):1697–1699.

[Tesauro, 1992] Tesauro, G. (1992). *Practical issues in temporal difference learning*. Springer.

[Tourbot et al., 2012] Tourbot, G., Bougerol, C., Glas, F., Zagonel, L. F., Mahfoud, Z., Meuret, S., Gilet, P., Kociak, M., Gayral, B., and Daudin, B. (2012). Growth mechanism and properties of ingan insertions in gan nanowires. *Nanotechnology*, 23(13):135703.

[Urban et al., 2015] Urban, A., MuÌĹller, M., Karbaum, C., Schmidt, G., Veit, P., Malindretos, J., Bertram, F., Christen, J., and Rizzi, A. (2015). Optical emission of individual gan nanocolumns analyzed with high spatial resolution. *Nano Lett.*, 15(8):5105–5109.

[Vampola et al., 2009] Vampola, K. J., Iza, M., Keller, S., DenBaars, S. P., and Nakamura, S. (2009). Measurement of electron overflow in 450 nm InGaN light-emitting diode structures. *Appl. Phys. Lett.*, 94(6):–.

[Vurgaftman et al., 2001] Vurgaftman, I., Meyer, J. R., and Ram-Mohan, L. R. (2001). Band parameters for iii-v compound semiconductors and their alloys. *J. Appl. Phys.*, 89(11):5815–5875.

[Wachutka, 1990] Wachutka, G. (1990). Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 9(11):1141–1149.

[Waibel et al., 1989] Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., and Lang, K. (1989). Phoneme recognition using time-delay neural networks. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(3):328–339.

[Waide et al., 2006] Waide, P., Tanishima, S., et al. (2006). *Light's Labour's Lost: Policies for Energy-efficient Lighting*. OECD Publishing.

[Wallace et al., 2014] Wallace, M. J., Edwards, P. R., Kappers, M. J., Hopkins, M. A., Oehler, F., Sivaraya, S., Allsopp, D. W. E., Oliver, R. A., Humphreys, C. J., and Martin, R. W. (2014). Bias dependence and correlation of the cathodoluminescence and electron beam induced current from an ingan/gan light emitting diode. *J. Appl. Phys.*, 116(3):033105.

[Wallace et al., 2015] Wallace, M. J., Edwards, P. R., Kappers, M. J., Hopkins, M. A., Oehler, F., Sivaraya, S., Oliver, R. A., Humphreys, C. J., Allsopp, D. W. E., and Martin, R. W. (2015). Effect of the barrier growth

mode on the luminescence and conductivity micron scale uniformity of ingan light emitting diodes. *J. Appl. Phys.*, 117(11):115705.

[Watson-Parris et al., 2011] Watson-Parris, D., Godfrey, M. J., Dawson, P., Oliver, R. A., Galtrey, M. J., Kappers, M. J., and Humphreys, C. J. (2011). Carrier localization mechanisms in $in_xga_{1-x}n$/gan quantum wells. *Phys. Rev. B*, 83:115321.

[Wiener, 1949] Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series*, volume 2. MIT press Cambridge, MA.

[Wu et al., 2002] Wu, L., Chang, S., Wen, T., Su, Y., Chen, J., Lai, W., Kuo, C.-H., Chen, C., and Sheu, J. K. (2002). Influence of si-doping on the characteristics of ingan-gan multiple quantum-well blue light emitting diodes. *Quantum Electronics, IEEE Journal of*, 38(5):446–450.

[Wyss and Booth, 1997] Wyss, M. and Booth, D. C. (1997). The iaspei procedure for the evaluation of earthquake precursors. *Geophys. J. Int.*, 131(3):423–424.

[Xie et al., 2008] Xie, J., Ni, X., Fan, Q., Shimada, R., Ozgur, U., and Morkoc, H. (2008). On the efficiency droop in InGaN multiple quantum well blue light emitting diodes and its reduction with p-doped quantum well barriers. *Appl. Phys. Lett.*, 93(12):121107.

[Yan et al., 1990] Yan, R.-H., Corzine, S., Coldren, L., and Suemune, I. (1990). Corrections to the expression for gain in gaas. *Quantum Electronics, IEEE Journal of*, 26(2):213–216.

[Zagonel et al., 2010] Zagonel, L. F., Mazzucco, S., Tencé, M., March, K., Bernard, R., Laslier, B., Jacopin, G., Tchernycheva, M., Rigutti, L., Julien, F. H., et al. (2010). Nanometer scale spectral imaging of quantum emitters in nanowires and its correlation to their atomically resolved structure. *Nano Lett.*, 11(2):568–573.

[Zagonel et al., 2012] Zagonel, L. F., Rigutti, L., Tchernycheva, M., Jacopin, G., Songmuang, R., and Kociak, M. (2012). Visualizing highly localized luminescence in gan/aln heterostructures in nanowires. *Nanotechnology*, 23(45):455205.

[Zhou et al., 2015] Zhou, X., Lu, M.-Y., Lu, Y.-J., Jones, E. J., Gwo, S., and Gradecak, S. (2015). Nanoscale optical properties of indium gallium nitride/gallium nitride nanodisk-in-rod heterostructures. *ACS Nano*, 9(3):2868–2875.